Rational Inattention with Ambiguity Aversion[*]

Francesco Fabbri[†]

June 24, 2024

Abstract

We develop a tractable framework to incorporate ambiguity aversion into rational inattention. We model uncertainty using smooth ambiguity (Klibanoff, Marinacci, and Mukerji, 2005) and define entropy-based information costs on the predictive prior distribution. This allows us to rewrite the problem in terms of stochastic choice rules. Our solution generalizes Matějka and McKay's (2015) multinomial logit formula by adding an exponential multiplicative state fixed-effect that depends on ambiguity and the attitudes towards it. We provide an axiomatic characterization of this formulation. We also study when our solution follows Luce's (1959) multinomial logit model, providing a new foundation robust to ambiguity aversion.

JEL Classification: D81, D83.
Keywords: Rational Inattention, Ambiguity, Multinomial Logit.

[*]I am grateful to Pietro Ortoleva for his guidance and encouragement. I also thank Pierpaolo Battigalli, Roland Benabou, Alex Bloedel, Sylvain Chassang, Mark Dean, Tommaso Denti, Satoshi Fukuda, Faruk Gul, Sam Kapon, Navin Kartik, Alessandro Lizzeri, Fabio Maccheroni, Massimo Marinacci, Yusufcan Masatlioglu, Xiaosheng Mu, Georgios Nikolakoudis, Wolfgang Pesendorfer, Luciano Pomatto, Giulio Principi, Doron Ravid, Nicola Rosaia, Kota Saito, Karthik Sastry, Chris Sims, Lorenzo Stanca, João Thereze, Leeat Yariv, and all seminar participants at Princeton University, D-Tea conference 2022, BRIC 2023 for their insightful questions, comments, and discussions.

[†]Princeton University, Department of Economics, ffabbri@princeton.edu. I acknowledge the financial support of the Bonaldo Stringher Scholarship by Bank of Italy and DETC grant.

1 Introduction

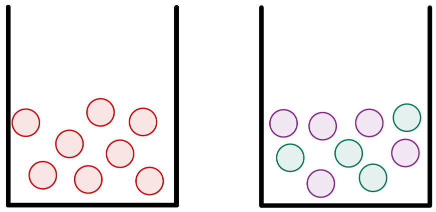
Models of rational inattention, where individuals trade-off between processing more information to improve decisions and saving on the mental effort, assume a well-defined prior probability over the possible states of the world. However, the most interesting cases of information processing arise when data is complex to analyze, e.g., demand's idiosyncratic conditions (Mackowiak and Wiederholt, 2009), portfolio choice with multiple assets (Mondria, 2010), or future values of risky assets (Van Nieuwerburgh and Veldkamp, 2010). These are precisely the situations where it is reasonable to expect that individuals may fail to have a well-defined belief over the possible states of the world and may instead be ambiguity averse. How does optimal information processing change if we allow for ambiguity averse preferences?

Our model enriches the rational inattention (RI) framework by incorporating ambiguity considerations in a tractable way. Before choosing, a decision maker (DM) costly processes information about an unknown payoff-relevant state, drawn according to an unknown probability distribution. We model uncertainty using smooth ambiguity (Klibanoff, Marinacci, and Mukerji, 2005), which allows us to distinguish between the degree of perceived ambiguity and attitudes towards it. This modeling approach allows us to define an entropy-based information processing cost in terms of the average prior distribution, which we refer to as *predictive*,[1] providing a way to generalize Sim's (2003) cost function to settings where multiple probability distributions, or models, are possible. Our first result (Lemma 1) shows we can simplify this information problem with both RI and ambiguity aversion by equivalently writing it in terms of stochastic choice rules, i.e., choice probabilities conditional on each state.

Illustrative example. Consider two urns: Urn 1 contains only red balls; Urn 2 an unknown proportion of green and purple balls. One ball is drawn randomly from one urn, and its color represents the payoff-relevant state. The DM can costly process information about the color drawn before choosing between RG, which yields a payoff of 1 if the drawn ball is either red or green, and P, which yields a payoff of 1 if purple.

First, assume that the individual follows expected utility. Notice that the states "red ball" and "green ball" are payoff-equivalent, as they yield the same payoff action-

[1]As the smooth ambiguity model ascribes two levels of uncertainty, i.e., the first on payoff-relevant states and the second on the possible models governing them, we can define a predictive prior distribution over states by averaging the weight assigned to each state by every model.



	red ball	green ball	purple ball
RG	1	1	0
P	0	0	1

Figure 1: Illustration of the problem with two urns.

wise. Since, under expected utility, information has only instrumental value, the DM would never waste cognitive resources to distinguish between them. As a result, the DM exhibits the same behavior conditional on both "red ball" and "green ball". More formally, this implies

$$\frac{\mathbb{P}(RG|\text{red ball})}{\mathbb{P}(\ P\ |\text{red ball})} = \frac{\mathbb{P}(RG|\text{green ball})}{\mathbb{P}(\ P\ |\text{green ball})}$$

meaning that the probability of betting on the *red or green* ball compared to the *purple* ball is unaffected by conditioning on the state "red ball" or "green ball". This property summarizes the content of Axiom 1 by Matějka and McKay (2015) and *invariance under compression* (IUC) by Caplin et al. (2022).

Adding ambiguity aversion changes these considerations. A DM that decides how much attention to pay to the state "green ball," reflected in the probability of choosing the optimal bet RG, should recognize that this state is related to "purple ball" through Urn 2, which is not the case for "red ball." That is, attention allocation serves as an instrument to hedge against the ambiguity induced by each state. As "red ball" and "green ball" induce different distributions on models, i.e., a Dirac measure on Urn 1 and Urn 2, respectively, they are not *information-equivalent*: An ambiguity averse DM may find it optimal to differentiate between them while processing information.

Main result. Our main result (Theorem 1) enriches Matějka and McKay's (2015) solution of the expected utility case by incorporating preferences for hedging against ambiguity. The behavior of a rational inattentive, ambiguity averse decision maker follows a biased multinomial logit model

$$\mathbb{P}(\text{action } a|\text{state } s) = \frac{\exp[p(a) + I(s) \cdot u(a,s)/\lambda]}{\sum_{b\in A} \exp[p(b) + I(s) \cdot u(b,s)/\lambda]} \tag{1}$$

where: $u(a,s)$ is the payoff associated with action a in state s; λ is the unit cost of information; p and I, which depends on ambiguity, capture fixed-effects associated with actions and states, respectively. In particular:

- The ratio $u(a,s)/\lambda$ represents the fundamental RI trade-off, stating that the probability of playing action a in state s is positively affected by the payoff associated with the action in that state but negatively by the information costs.
- The function p, as in Matějka and McKay (2015), represents the *a priori* probability of choosing each action. Specifically, if an action is a priori more appealing, it will be chosen more often in each state.
- The function $I \geqslant 0$ is novel and pertains to ambiguity aversion. It measures the states' *information value* by evaluating the information each provides about possible models. Notably, in the absence of ambiguity or with neutral attitudes towards it, I reduces to 1, and equation (1) collapses to the Matějka and McKay's (2015) solution. For every state s with $I(s) \geqslant 1$, we can view I as amplifying the value of the ratio $I(s) \cdot u(\cdot, s)/\lambda$ relatively to the expected utility case, positively impacting the attention paid at s. Therefore, ambiguity aversion influences through I the incentives to pay attention to specific states by modifying their rewards.

Next, we characterize (Propositions 1 and 2) the states that receive more (less) attention as the ones associated with models that yield lower (higher) expected utilities. This result implies that ambiguity averse DMs manipulate their attention to hedge against uncertainty, re-balancing the information process by allocating more attention in favor of states associated with lower expected utility levels. Furthermore, we show (Corollary 2) that if one state has a higher information value than another state, signals about it are more precise.

Foundations. We discuss two behavioral implications of our model. First, as illustrated by the example above, adding ambiguity aversion leads to violations of invariance under compression (IUC).[2] However, we show (Proposition 3) that it satisfies a weaker form that requires "compression" for states that are both payoff and information-equivalent, that is, have the same information value. Second, we prove

[2]This implies that the result by Caplin et al. (2022) stating that IUC characterizes entropy-based information costs in the larger class of posterior separable costs is not robust to ambiguity aversion. Therefore, IUC uniquely identifies entropy-based information under expected utility.

(Proposition 6) that RI with ambiguity aversion is unaffected by Debreu's (1960) criticism of Luce's (1959) multinomial logit model, which states that adding a duplicate action can affect the choice probabilities of non-duplicate actions. This result conforms with the one of Matějka and McKay (2015) for the expected utility case.

We investigate the conditions under which our solution (1) collapses to Luce's (1959). Under expected utility, Matějka and McKay (2015) show that this happens when actions are *a priori homogeneous*, i.e., ex-ante indistinguishable payoff-wise. This is no longer sufficient under ambiguity aversion, as we further need to restrict states' information value, which influences their associated rewards.[3] However, under an additional condition restricting the set of possible models, this influence occurs homogeneously, and we show (Proposition 4) that the solution of RI with ambiguity aversion follows Luce's multinomial logit model. As we do not restrict preferences for uncertainty, this result provides a micro-foundation that is robust to any ambiguity aversion level.[4] We also show (Corollary 1) that the same result holds for maxmin preferences (Gilboa and Schmeidler, 1989).

Next, we provide an axiomatic foundation for the solution (1). As discussed above, our model violates Axiom 1, one of the two axioms of Matějka and McKay (2015), implying that, also under RI, ambiguity averse and expected utility behavior are distinguishable. To assess whether a DM's choices are compatible with RI for some ambiguity aversion level, we take the perspective of an analyst who constructs statistics using the available data, composed solely of stochastic choice rules, to represent the elements p, u, and I in equation (1). We characterize (Theorem 2) the properties stochastic rules must satisfy to allow this construction. In particular, we generalize Axiom 1 of Matějka and McKay (2015) to account for ambiguity averse behavior, and keep their Axiom 2, that requires duplicates to be treated as a single action.

We discuss additional results related to our model. First, by restricting ambiguity attitudes to constant absolute ambiguity aversion (CARA), we show (Proposition 5) that the average of the information values of each state equals 1, implying that not all states can amplify or dampen their associated rewards compared to the expected utility case. Second, we modify the information processing domain, allowing the DM to pay costly attention to models and not states. We show that this framework can be represented as RI under expected utility.

[3] Under ambiguity aversion, the reward associated with action a in state s is $I(s) \cdot u(a, s)$, where $I(s)$ may differ from one.

[4] Assumption 1 restricts ambiguity attitudes to ensure that the optimization problem (2) faced by the DM is convex. Our result is robust to any ambiguity aversion level that satisfies this assumption.

Related Literature. We contribute directly to the literature on rational inattention initiated by Sims (2003). This paper generalizes the model of Matějka and McKay (2015) by incorporating ambiguity aversion: Our solution collapses to theirs if the DM is ambiguity neutral or only one model is possible. We show how to strengthen their condition of *a priori homogeneity* to obtain Luce's (1959) multinomial logit model for any ambiguity aversion level. Furthermore, we weaken their Axiom 1 to characterize rationally inattentive behavior under ambiguity aversion.

The concurrent work of Hansen et al. (2022) studies RI with model misspecification. However, our approach differs from theirs as they study robustness by assuming variational preferences (Maccheroni, Marinacci, and Rustichini, 2006), while we incorporate ambiguity attitudes using the smooth model by Klibanoff et al. (2005). This difference has substantive implications: As smooth ambiguity quantifies the uncertainty over models, we can define an entropy-based information cost on the predictive prior distribution, making the cost independent from preferences. As a result, the solution of Hansen et al. (2022) follows the multinomial logit formula by Matějka and McKay (2015) according to some *worst-case* prior, while our solution differs from it by adding an exponential multiplicative state-fixed effect. At the behavioral level, this difference has bite as our solution implies a violation of Axiom 1, which is instead satisfied in their case for the worst-case prior. Besides this difference, their study complements ours by discussing different results concerning RI and robustness.

We explore the assumption of *invariance under compression* (IUC), introduced by Caplin et al. (2022). In particular, we show that an ambiguity averse DM violates it unless the compression occurs for both payoff and information-equivalent states, a property that we call *weak invariance under compression.* In experimental tests of rational inattention, Dean and Neligh (2023) show that subjects violate IUC, suggesting that some states are harder to investigate than others. Our model relates state-dependent information costs to ambiguity aversion. Other relevant works are Caplin and Dean (2013) and Caplin et al. (2019), which, as in our axiomatic characterization, tackle RI in terms of induced posterior distributions.

One of our results allows us to rewrite RI with smooth ambiguity in terms of stochastic choice rules, i.e., maps from states to distribution over actions, which we can interpret as *acts* in the standard decision theory sense. For this reason, our ambiguity preferences can be represented by a static model, which contributes to making our analysis tractable, as dynamic ambiguity models suffer either from dynamic inconsistencies or inconsequential updating (Epstein and Schneider, 2003; Hanany and Klibanoff, 2007, 2009; Siniscalchi, 2011).

Finally, our work fits into the literature on ambiguity and information. Li and Zhou (2016) shows that our preferences satisfy the Blackwell order. Auster et al. (2024) and Epstein and Ji (2022) investigate dynamic information acquisition under ambiguity. Differently from us, they assume that information costs are not entropy-based but time-dependent. Lu (2021) studies random choices resulting from exogenous shocks to information and ambiguity aversion. We also study random choice under ambiguity aversion, but our source of stochasticity is endogenous information processing. Denti and Pomatto (2022) provide necessary and sufficient conditions for smooth ambiguity preferences to be statistically identifiable. One of our conditions in Proposition 1 meets their requirements. Mele and Sangiorgi (2015) apply information acquisition with ambiguity to study asset markets. Battigalli et al. (2015) study the effect of ambiguity on experimentation in the context of self-confirming equilibrium. Other related works at this intersection are Li (2020) and De Oliveira et al. (2021).

2 Model

Before choosing an action from a finite set A, the decision maker (DM) can costly process information about an unknown payoff-relevant state $\theta \in \Theta$, which is drawn from a finite set according to the unknown probability distribution $\pi \in \Delta(\Theta)$. We refer to elements in $\Delta(\Theta)$ as *models* and allow the DM to form a *prior distribution* on them, $\mu \in \Delta(\Delta(\Theta))$ with $\Pi := \operatorname{supp} \mu$. For every state $\theta \in \Theta$, define the *predictive distribution* as $\rho(\theta) := \mathbb{E}_{\mu}[\pi(\theta)]$. It captures the expected probability of each state and plays a key role in our model as we use it to compute the information costs. Assume ρ has full support, a condition that holds if, for every $\theta \in \Theta$, there exists a $\pi \in \Pi$ such that $\pi(\theta) > 0$. Denote by X an arbitrary finite set of signals.[5]

An *attention strategy* $\sigma = (F, \alpha)$ is a pair composed of:

1. An *information structure* $F : \Theta \to \Delta(X)$ that associates each state with a distribution of signals,[6]

[5]The restriction to finite actions, signals, and states is convenient as it allows us to avoid measurability issues and ease the exposition. Our results extend to standard continuous models following the techniques of Csiszár (1974) and Denti et al. (2020).

[6]Even if we allow for arbitrary (finite) signal spaces, the information structure is still constrained by its domain. That is, the DM can process information about payoff-relevant states but not about other features, such as the models. For example, in the two urns illustration in the introduction, the DM can consult statistics to assess the color of the drawn ball but cannot ask for additional *i.i.d.* draws, as they provide information about the urn composition, and not about the state.

2. An *action plan* $\alpha : X \to A$ that denotes which action to implement after each signal realization.[7]

Following the literature on rational inattention initiated by Sims (2003), we assume that information processing costs are entropy-based. The entropy of a probability distribution $p \in \Delta(S)$, defined over a finite set S, is

$$H(p) := \sum_{s \in S} p(s) \log p(s).$$

The cost associated with the information structure $F : \Theta \to \Delta(X)$ is proportional to the average reduction in entropy between unconditional and conditional predictive distributions

$$c(\rho, F) := \lambda \left(H(\rho) - \mathbb{E}_x[H(\rho(\cdot|x))] \right),$$

where $\lambda \geqslant 0$ is the unit information cost, and $\rho(\cdot|x) \in \Delta(\Theta)$ denotes the *posterior predictive distribution* associated with signal $x \in X$. That is, processing information F is more costly if the difference between the uncertainty *ex-ante* and *ex-post*, measured respectively as the entropy of the predictive distribution $H(\rho)$ and the average entropy of each predictive posterior $H(\rho(\cdot|x))$, is larger.

The DM may feature ambiguity aversion when multiple models are possible, i.e., $|\Pi| > 1$. Each action $a \in A$ is evaluated according to the smooth ambiguity certainty equivalent, expressed in utils, by Klibanoff et al. (2005)

$$V^{(\phi,\mu)}(a) := \phi^{-1} \left(\mathbb{E}_\mu \left[\phi \left(\sum_{\theta \in \Theta} u(a, \theta) \pi(\theta) \right) \right] \right)$$

where $u : A \times \Theta \to \mathbb{R}$ is the utility function, and $\phi : \overline{\text{co}}\, u(A \times \Theta) \to \mathbb{R}$ is a strictly increasing, concave, and smooth function that represents the DM's ambiguity attitudes.[8] Recall that linearity of ϕ implies ambiguity neutrality while strict concavity implies ambiguity aversion.

[7]Choosing an action plan amounts to commit to an action for each possible signal realization, ensuring the ex-ante value of the information problem. As non-expected utility preferences may fail to be dynamically consistent, the ex-ante and ex-post information values may differ. For instance, in the absence of commitment, an ambiguity averse DM can reject freely available information, as shown by Siniscalchi (2011). The commitment assumption allows us to study the benefits of ex-ante information processing freed from the considerations concerning dynamic inconsistency.

[8]Here, $\overline{\text{co}}\, u(A \times \Theta)$ is the smallest closed interval containing the image of the function u, and smooth means at least four times differentiable.

Definition 1. The rational inattention problem with smooth ambiguity is

$$\max_{(F,\alpha)} \phi^{-1}\left(\mathbb{E}_\mu\left[\phi\left(\sum_{\theta\in\Theta}\sum_{x\in X} u(\alpha(x),\theta)\, F(x|\theta)\, \pi(\theta)\right)\right]\right) - c(\rho, F). \tag{2}$$

To sum up, the DM chooses an information structure F that affects the entropy costs and the expected utility of every model $\pi \in \Pi$, calculated following the action plan $\alpha : X \to A$. The DM evaluates these expectations with the smooth function ϕ, which then aggregates using the prior μ. Finally, the DM normalizes the ambiguous utils obtained by ϕ^{-1}. Define this certainty equivalent as $V^{(\phi,\mu)}(F, \alpha)$.

Definition 1 subsumes two relevant modeling assumptions:

(i) Information costs do not depend on ambiguity attitudes and are affected by the prior over models, $\mu \in \Delta(\Pi)$, only through the predictive distribution, $\rho \in \Delta(\Theta)$. This implies that the information processing cost is *ambiguity-less*, capturing the idea that ambiguity averse and expected utility agents face identical entropy costs as long as they share the same predictive distribution.

(ii) Information processed about payoff-relevant states, $F : \Theta \to \Delta(X)$, does not update the prior over models, guaranteeing that no model can be discarded as a result. This, together with the commitment to a signal-contingent action plan, ascribes $V^{(\phi,\mu)}$ to the class of certainty equivalents studied by Li and Zhou (2016) that satisfy the Blackwell's informativeness ranking over information structure.[9] If this property failed, utility and attention incentives would be misaligned: Processing Blackwell's more valuable information would cost more in terms of entropy without providing any utility benefit, which conflicts with the interpretation of entropy-based costs being cognitive.

Furthermore, as formalized by Lemma 1, the combination of commitment to an action plan, points (i) and (ii) above, improves the tractability of the problem, by, respectively, reducing the information decision to a static one, simplifying how uncertainty resolves, and allowing an equivalent formulation in terms of stochastic choice rules.

[9] The information structure F is Blackwell more informative than G if there exists a stochastic kernel $T : X \to \Delta(X)$ such that $G = FT$. By Li and Zhou (2016), we have that F is Blackwell more informative that G if and only if $V^{(\phi,\mu)}(F, \alpha) \geqslant V^{(\phi,\mu)}(G, \alpha)$ for every α, μ, and concave ϕ.

2.1 Preliminary Results

As a first step, we restate RI with smooth ambiguity (2) in terms of conditional choice probabilities rather than attention strategies. Intuitively, the choice of an information structure suffices to describe a solution to the problem by identifying each signal with the "recommendation" to play the corresponding optimal action. Formally, we introduce a *stochastic choice rule* $f : \Theta \to \Delta(A)$ which assigns a distribution over actions, denoted by $f(\cdot|\theta)$, to each state $\theta \in \Theta$. A strategy $\sigma = (F, \alpha)$ generates the stochastic choice rule f if, for every $a \in A$ and $\theta \in \Theta$,

$$f(a|\theta) = \mathbb{P}(\{x \in X : \alpha(x) = a\}|\theta) = \sum_{x \in X_a} F(x|\theta)$$

where $X_a := \{x \in X : \alpha(x) = a\}$ is the set of signals associated with action a by the action plan α. Conversely, every stochastic choice rule f induces a strategy (F, α) by identifying each action with a signal realization in X and then equating the probability of each signal realization to the probability of the corresponding action.

We denote by f^0 the *expected unconditional* probability of playing action a

$$f^0(a) := \sum_{\theta \in \Theta} f(a|\theta)\rho(\theta).$$

Furthermore, the stochastic choice rule f associates each model $\pi \in \Pi$ with the expected utility

$$U(\pi, f) := \sum_{\theta \in \Theta} \sum_{a \in A} u(a, \theta) f(a|\theta) \pi(\theta).$$

The following result, which generalizes an equivalent one under expected utility in Matějka and McKay (2015), shows that the optimal strategy and its generated stochastic choice rule induce the same payoff to the DM. This implies that it is without loss to focus on stochastic choice rules to characterize the optimal solution.

Lemma 1. *If the attention strategy σ solves RI with smooth ambiguity* (2)*, then the corresponding stochastic choice rule f solves*

$$\max_f \phi^{-1}\left(\mathbb{E}_\mu\left[\phi\left(U(\pi, f)\right)\right]\right) - c(f, \rho) =: V(f) - c(f, \rho). \tag{3}$$

Conversely, any stochastic choice rule that solves (3) *induces an attention strategy that solves RI with smooth ambiguity* (2)*.*

The proof of Lemma 1 shows, by the convexity of entropy-based costs, that optimal information processing never induces different posteriors, obtained by updating the predictive distribution, conditional on signals recommending the same action, i.e., for every $x_1, x_2 \in X_a$, we have $\rho(\cdot|x_1) = \rho(\cdot|x_2)$. This establishes a one-to-one relationship between actions and predictive posteriors, and the lemma follows by the symmetry of the cost function.

Equation (3) sheds light on the static nature of our model, an implication of commitment to an action plan and the previously introduced assumptions (i) and (ii). We can interpret any stochastic choice rule $f : \Theta \to \Delta(A)$ as an *act*, that is, a function that maps every state to a probability distribution over a finite set of *consequences* A. Under this interpretation, $V(f)$ is the value that the smooth ambiguity model with parameters (μ, ϕ) associates to act f. Furthermore, we can see $f^0 \in \Delta(A)$ as a *constant act*, describing the ex-ante expected consequences of implementing act f. The cost

$$c(f, \rho) := \lambda \left(H(f^0) - \mathbb{E}_\rho[H(f(\cdot|\theta))] \right)$$

penalizes behavior diversification, by measuring, in entropy terms, the extent to which the act f induces, on average, a different distribution of consequences for every state. The following example illustrates this interpretation.

Example 1. Consider the following payoff structure

	θ_1	θ_2
a	1	0
b	0	1

where the rows are actions, the columns are states, and the matrix entries represent the payoffs. Let $\rho = 1/2$. Define the acts $f_1, f_2 : \Theta \to \Delta(A)$ as follows

$$f_1(a|\theta_1) = f_1(b|\theta_2) = 1 - \varepsilon$$
$$f_2(a|\theta_1) = f_2(b|\theta_2) = 1/2$$

for $\varepsilon \in [0, 1/2)$. In words, f_1 prescribes the optimal action in each state with probability $1 - \varepsilon$, while f_2 randomizes uniformly. By monotonicity, any ambiguity averse DM that can diversify behavior across states without incurring any cost, i.e., $\lambda = 0$, prefers f_1 over f_2. However, if $\lambda > 0$ is sufficiently high, distinguishing the states may be too costly and f_2 may be preferred.

To see this, notice that the acts f_1 and f_2 induce the same a priori constant act, i.e., $f_1^0 = f_2^0 = 1/2$, which coincides with f_2. Therefore, the act f_2 is costless in entropy terms. On the other hand, the cost of the act f_1 is:

$$c(f_1, \rho) = \lambda \big(\log(1/2) - (\varepsilon \log(1-\varepsilon) + (1-\varepsilon)\log(\varepsilon))\big) > 0 = c(f_2, \rho)$$

for every $\lambda > 0$. Therefore, for every smooth parameter (μ, ϕ), there exists a threshold of unit cost $\lambda > 0$ such that f_2 is preferred over f_1. ◂

One issue with RI with smooth ambiguity (3) is that it may not constitute a convex optimization problem, even under ambiguity aversion. We circumvent this issue by restricting the ambiguity attitudes. The following assumption, due to Hardy et al. (1934) and Hennessy and Lapan (2006), characterizes concave certainty equivalents. We require it holds throughout the remainder of the paper.

Assumption 1. The reciprocal of the Arrow-Pratt coefficient of absolute ambiguity aversion, $1/r(v) := -\phi'(v)/\phi''(v)$, where $v \in \overline{\text{co}}\, u(A \times \Theta)$, is concave.

While Assumption 1 is stronger than requiring strict concavity of ϕ, it is satisfied by most preferences used in applications. In particular, it holds for constant absolute ambiguity aversion (CARA), which nests variational preferences (Maccheroni et al., 2006), and constant relative ambiguity aversion (CRRA).[10]

Lemma 2. *If Assumption 1 is satisfied, then RI with smooth ambiguity* (3) *is a convex optimization problem.*

Lemma 2 implies the Karush-Kuhn-Tucker conditions are necessary for optimality, while the finiteness of A and Θ ensures the existence of an optimal solution. However, uniqueness is not always guaranteed, as a trivial multiplicity may arise due to redundancies in the payoff structure.[11]

[10] If $1/r(v)$ is linear, and hence Assumption 1 holds, then ϕ satisfies hyperbolic absolute ambiguity aversion (HARA),

$$\phi(x) = \frac{1-\gamma}{\gamma}\left(\frac{\kappa x}{1-\gamma} + \eta\right)^{\gamma}$$

for $\kappa > 0$, $\frac{\kappa x}{1-\gamma} + \eta > 0$. CARA is obtained when γ goes to infinity, CRRA when $\eta = 0$.

[11] See the paragraph on duplicate actions in section 5 or Matějka and McKay (2015) for a discussion.

3 Solution

Our main result characterizes the stochastic choice rule that solves RI with smooth ambiguity (3). The solution extends the multinomial logit formulation of Matějka and McKay (2015) by incorporating ambiguity considerations.

Theorem 1. *If $\lambda > 0$, the stochastic choice rule f that solves RI with smooth ambiguity* (3) *satisfies*

$$f(a|\theta) = \frac{f^0(a) \exp[u(a,\theta) \cdot \mathcal{I}_\theta(f)/\lambda]}{\sum_{b \in A} f^0(b) \exp[u(b,\theta) \cdot \mathcal{I}_\theta(f)/\lambda]} \tag{4}$$

where

$$\mathcal{I}_\theta(f) := \frac{\mathbb{E}_{\mu_\theta}[\phi'(U(\pi, f))]}{\phi'(V(f))} \tag{5}$$

and $\mu_\theta(\cdot) := \mu(\cdot|\theta) \in \Delta(\Pi)$. If $\lambda = 0$, then in each state the actions that yield the highest utility are chosen with probability one.

Theorem 1 generalizes the RI decision problem allowing for non-neutral ambiguity attitudes. If only one model is possible or if the DM is ambiguity neutral, i.e., $|\Pi| = 1$ or ϕ is linear, respectively, then $\mathcal{I}_\theta(\cdot) = 1$ for every state $\theta \in \Theta$.[12] Thus, equation (4) simplifies to

$$f(a|\theta) = \frac{f^0(a) \exp[u(a,\theta)/\lambda]}{\sum_{b \in A} f^0(b) \exp[u(b,\theta)/\lambda]} \tag{6}$$

which is the solution found by Matějka and McKay (2015). In this case, for every action $a \in A$ and state $\theta \in \Theta$, $f(a|\theta)$ follows an extended multinomial logit. The DM plays more often the actions that are *a priori* more appealing, as captured by $f^0(a)$. The exponential term is evaluated at $u(a,\theta)/\lambda$, which constitutes the RI trade-off between utility gains and information costs.

Adding ambiguity aversion enriches the standard solution by multiplying the argument in the exponential term by $\mathcal{I}_\theta(f) \geqslant 0$. Following equation (5), $\mathcal{I}_\theta(f)$ can be interpreted as the *information value* of state $\theta \in \Theta$ given the stochastic choice rule f. The expectation at the numerator of equation (5) is taken with respect to the probability distribution $\mu_\theta \in \Delta(\Pi)$, obtained by updating the prior $\mu \in \Delta(\Pi)$

[12] If $\Pi = \{\pi\}$, $\mathbb{E}_{\mu_\theta}[\phi'(U(\pi, f))] = \phi'(U(\pi, f)) = \phi'(V(f))$, and if ϕ is linear, $\phi' = k$ for some $k \in \mathbb{R}$.

conditional on the state $\theta \in \Theta$,[13] which represents DM's ambiguity perception before choosing $f(\cdot|\theta) \in \Delta(A)$. The denominator of (5) is constant across states and serves as a normalization. Therefore, if ϕ is concave, then ϕ' is decreasing, which implies that we have a lower (higher) value of $\mathcal{I}_\theta(f)$ when μ gives higher (lower) weight to models associated with high expected utility. As a result, $\mathcal{I}_\theta(f)$ modifies the standard RI trade-off for state θ, by increasing (decreasing) the attentional reward $\tilde{u}(\cdot,\theta) := u(\cdot,\theta) \cdot \mathcal{I}_\theta(f)$ if μ_θ places more (less) weight on models featuring lower (higher) levels of expected utility. The desire to hedge against ambiguity affects information processing by providing incentives to pay more (less) attention to states associated with lower (higher) expected utility levels.

We say that state $\theta \in \Theta$ is of *high information value* if $\mathcal{I}_\theta \geqslant 1$, and of *low information value* otherwise. This captures the idea that if a state has a high (low) information value, then any ambiguity averse DM has stronger (weaker) incentives to pay attention to it compared to an expected utility agent, i.e., $\tilde{u}(a,\theta) \geqslant u(a,\theta)$ ($\tilde{u}(a,\theta) \leqslant u(a,\theta)$) for every action $a \in A$. The following result formalizes the previous discussion by characterizing states with high information value in terms of their associated expected payoff. As it relies on the curvature of ϕ', recall that ϕ is *prudent* when ϕ' is convex, and *imprudent* when ϕ' is concave.

Proposition 1. *Let $\theta \in \Theta$, and f be a stochastic choice rule.*

(i) $\mathcal{I}_\theta(f) \geqslant 1$*, i.e., the state θ has high information value,*

(ii) $\mathbb{E}_{\mu_\theta}[U(\pi, f)] \leqslant V(f)$.

If ϕ is prudent, (ii) implies (i), while if ϕ is imprudent (i) implies (ii). Furthermore, the two statements are equivalent if the support of μ_θ is a singleton for every $\theta \in \Theta$.

Point *(ii)* reads that the expected utility of the problem calculated using the distribution over models $\mu_\theta \in \Delta(\Pi)$ induced by the state $\theta \in \Theta$ is lower than the corresponding certainty equivalent. It suggests that $\theta \in \Theta$ is a "bad state" as it ex-post reduces the ex-ante value of the problem.

When the support of μ_θ is a singleton for every state $\theta \in \Theta$,[14] characterizing high information value states is relatively straightforward. In this case, the equivalence

[13]Formally, $\mu_\theta(d\pi) := \mu(d\pi|\theta) = \frac{\pi(\theta)}{\rho(\theta)}\mu(d\pi)$ for every $\pi \in \Pi$, $\theta \in \Theta$.

[14]Although restrictive, this condition is met by the identifiable smooth model by Denti and Pomatto (2022).

between (i) and (ii) formalizes the idea that states with high (low) information value are associated with models that yield lower (higher) expected utility levels. However, in more general cases, this relation depends on the curvature of ϕ'. In particular, if ϕ' is prudent, then any "bad state" θ is associated with a high information value, thus increasing $\tilde{u}(\cdot,\theta)$ compared to the expected utility case. It is still possible that some "good state" receives the same attentional reward. On the other hand, if ϕ' is imprudent then every "good state" θ, defined as $\mathbb{E}_{\mu_\theta}[U(\pi, f)] \geqslant V(f)$, is associated with low information value, but also some "bad state" may. Interestingly, point (i) and (ii) are again equivalent if ϕ is both prudent and imprudent, i.e., ϕ' is linear.

Next, we formalize the precise sense in which ambiguity aversion induces the DM to pay more attention to states with high information value. Stating the result requires additional notation. Denote by f^ϕ and $f^{0,\phi}$ the stochastic choice rule and unconditional probability over actions, respectively, that solve problem (3) when ambiguity attitudes are ϕ. For every action $a \in A$, the corresponding predictive posterior $\rho^\phi(\cdot|a) \in \Delta(\Theta)$, that is, the probability distribution over states given the "recommendation" to play a, is obtained by Bayes rule

$$\rho^\phi(\theta|a) := \frac{f^\phi(a|\theta)\rho(\theta)}{f^{0,\phi}(a)}.$$

Notice that the predictive posterior under ambiguity neutrality ρ^{id} is equivalent to the one obtained in the expected utility case.

Proposition 2. *Let $\theta \in \Theta$, and f^ϕ be a stochastic choice rule associated with ambiguity attitudes ϕ. The following are equivalent:*

(i) $\mathcal{I}_\theta(f^\phi) \geqslant 1$,

(ii) *For every action $a, b \in A$ with $u(a,\theta) \geqslant u(b,\theta)$ we have that*

$$\frac{\rho^\phi(\theta|a)}{\rho^\phi(\theta|b)} \geqslant \frac{\rho^{id}(\theta|a)}{\rho^{id}(\theta|b)} \geqslant 1.$$

We can separate point (ii) into two statements. The first says that, as for the expected utility case, if action $a \in A$ is preferred over action $b \in B$ at state $\theta \in \Theta$, then any ambiguity averse DM processes information such that the "recommendation" to play a is more informative about θ than the "recommendation" to play b. The second says that the informativeness of the "recommendation" to play a compared to b is higher under ambiguity aversion than expected utility.

The equivalence of (i) and (ii) implies that an ambiguity averse DM pays more (less) attention to states with high (low) information value by generating "signals" that induce more (less) favorable posterior distributions than under expected utility.

To show Proposition 2, we apply Theorem 1 to obtain

$$\frac{\rho^{\phi}(\theta|a)}{\rho^{\phi}(\theta|b)} = \left(\frac{\exp[u(a,\theta)/\lambda]}{\exp[u(b,\theta)/\lambda]}\right)^{\mathcal{I}_{\theta}(f^{\phi})} \tag{7}$$

which expresses the posterior ratio at state $\theta \in \Theta$, conditional on the "recommendations" to play $a, b \in A$, in terms of the corresponding payoff incentives, attention costs, and information value. Equation (7) generalizes the corresponding one under expected utility by allowing $\mathcal{I}_{\theta}(\cdot) \neq 1$.

Together, Proposition 1 and 2 formalize how information processing serves to hedge against ambiguity. They imply that any prudent (imprudent) ambiguity averse DM pays, in the sense of Proposition 2, more (less) attention to states associated with low (high) expected utility levels.

3.1 Invariance Under Compression

Caplin et al. (2022) show that the behavior of a DM with entropy-based information cost satisfies *invariance under compression* (IUC). Roughly, this axiom requires that if two states are payoff-equivalent, that is, they yield the same payoff for every action, it is never optimal to waste attention trying to distinguish them. They also show that, under expected utility, the entropy-based information cost is the only specification within the class of uniformly posterior separable cost functions[15] that satisfies IUC. Compression of payoff-equivalent states may not be optimal under ambiguity aversion, as the information value of each state additionally influences information processing.

The states $\theta_1, \theta_2 \in \Theta$ are *payoff-equivalent* if $u(a, \theta_1) = u(a, \theta_2)$ for every $a \in A$. We capture the essence of invariance under compression with the following definition.

Axiom (IUC)**.** *If $\theta_1, \theta_2 \in \Theta$ are payoff-equivalent, then $f(\cdot|\theta_1) = f(\cdot|\theta_2)$.*

IUC says that DM's behavior is invariant to payoff-equivalent states, which implies that choices are unaffected by changes in the state space that do not impact payoffs. However, as the following example illustrates, IUC is not satisfied under ambiguity.

[15] An information cost function k is *uniformly posterior separable* if

$$k(\rho, F) = \lambda\left(G(\rho) - \mathbb{E}_x[G(\rho(\cdot|x))]\right)$$

for some strictly concave and continuous $G : \Delta(\Theta) \to \mathbb{R}$. If k is entropy-based then $G = \mathrm{H}$.

Example 2. Consider the following payoff structure.

	θ_1	θ_2	θ_3
a	3	3	-y
b	2	2	0

Let $y > 0$ be such that $f^0(a), f^0(b) > 0$, that is, both actions are played with positive probability. The collection of possible models is $\{\pi_1, \pi_2\}$ where $\pi_1 = (1, 0, 0)$ and $\pi_2 = (0,\ 1/4,\ 3/4)$; the prior is $\mu = (1/3,\ 2/3)$.

Notice that, similarly to the illustrative example in the introduction, the model π_1 only predicts the state θ_1, while the model π_2 both states θ_2 and θ_3. Upon deciding the optimal plan of action conditional on state θ_2, the agent correctly infers that π_2 is the true model and, therefore, that the state θ_3 may occur as well. This argument does not apply to planning conditional on state θ_1, generating incentives to distinguish the payoff-equivalent states θ_1 and θ_2 while processing information. For instance, planning at θ_1 is unaffected by the value of y, while it affects planning at θ_2: The higher the value of y, the costlier the mistake of choosing a when the state is θ_3.

Formally, since the states θ_1 and θ_2 are payoff-equivalent, IUC requires that $f(a|\theta_1) = f(a|\theta_2)$. But, following equation (4), this is not the case since

$$\mu_{\theta_1} = \delta_{\pi_1} \neq \delta_{\pi_2} = \mu_{\theta_2}$$

implies $\mathcal{I}_{\theta_1} \neq \mathcal{I}_{\theta_2}$.[16] ◂

The example illustrates how adherence to IUC is not a consequence of the entropy-based cost alone but a joint property of rational inattention and expected utility. This joint property yields substantive economic consequences: Hébert and La′O (2023) shows it guarantees equilibrium existence with zero non-fundamental volatility in a class of generalized beauty contest games with endogenous information; Angeletos and Sastry (2024) shows it is necessary and sufficient to obtain a version of the First Welfare theorem for general equilibrium economies with inattentive agents. As the arguments of these papers rely on IUC, Example 2 suggests that their conclusions may not hold under ambiguity aversion.

[16] By contradiction, let $\mathcal{I}_{\theta_1} = \mathcal{I}_{\theta_2}$. By Proposition 3, this implies that $f(a|\theta_1) = f(a|\theta_2)$. For some $y > 0$, $f^0(a) > 0$ and, in particular, $f(a|\theta_1) > 0$. By equation (4), this implies that $f(a|\theta_3) > 0$. Now, $\mathcal{I}_{\theta_1} = \mathcal{I}_{\theta_2}$ is equivalent to, after some calculations, $f(a|\theta_1)u(a, \theta_1) + f(b|\theta_1)u(b, \theta_1) = f(a|\theta_3)u(a, \theta_3)$, which is not possible as the LHS is positive, while the RHS is negative.

The intuition for this failure is as follows. On the one hand, the entropy-based cost is *invariant*, i.e., roughly, the relabelling of states does not alter the information cost, and compressing states has no consequences as long as the signals do not distinguish between them.[17] On the other, an expected utility DM is indifferent between information structures that differ solely for the information provided about payoff-equivalent states, implying it is never optimal to costly distinguish them. This no longer holds under ambiguity aversion: DMs may find it optimal to differentiate payoff-equivalent states as long as they provide different information about models. This discussion suggests a weaker form of invariance, where we compress payoff and information-equivalent states; we formalize it next.

Two states $\theta_1,\ \theta_2 \in \Theta$ are *information-equivalent* if $\mu_{\theta_1} = \mu_{\theta_2}$. Notice that the notion of information-equivalence does not rely on the problem solution but only on DM's uncertainty perception. Furthermore, by equation (5), it is immediate to verify that, if $\theta_1, \theta_2 \in \Theta$ are information-equivalent, then $\mathcal{I}_{\theta_1} = \mathcal{I}_{\theta_2}$, conforming with our previous interpretation of information value.

Axiom (weak IUC)**.** *If $\theta_1,\ \theta_2 \in \Theta$ are payoff and information-equivalent, then $f(\cdot|\theta_1) = f(\cdot|\theta_2)$.*

Proposition 3. *If f solves RI with smooth ambiguity* (3)*, it satisfies weak IUC.*

Proposition 3 states that the RI solution is unchanged if we compress both payoff and information-equivalent states, implying that any ambiguity averse DM never distinguishes between states that share the same payoff structure and induce the same information about models. Notice that, since only one model is possible under expected utility, i.e., $|\Pi| = 1$, every state is information-equivalent, which implies that weak IUC collapses to IUC.

3.2 Multinomial Logit: Robust Foundation

We investigate when our solution reduces to Luce's (1959) multinomial logit model. For the expected utility case, Matějka and McKay (2015) show that Luce's model is obtained if every action is indistinguishable before processing information, a condition they call *a priori homogeneity*. In this section, we align with their framework by

[17] For a formal discussion on invariant information costs, see Hébert and La'O (2023), Angeletos and Sastry (2024); this notion originates from information geometry, Amari and Nagaoka (2000).

identifying states with the induced payoff-vector.[18] Furthermore, as duplicate actions are treated as a single one (Proposition 6), we assume no duplicates without loss.

Let $\tilde{\rho} \in \Delta(\mathbb{R}^A)$ be the probability over payoff-vectors defined as $\tilde{\rho}(v_1, \ldots, v_{|A|}) = \rho(\theta)$ if $v := (v_1, \ldots, v_{|A|}) = (u(a_1, \theta), \ldots, u(a_{|A|}, \theta))$ for $\theta \in \Theta$. *A priori homogeneity* holds if $\tilde{\rho}$ is invariant to all permutations of the indexes in v. Two states $\theta_1, \theta_2 \in \Theta$ are *exchangeable* if there exists an action-index permutation $\iota : |A| \to |A|$ such that $(u(a_1, \theta_1), \ldots, u(a_{|A|}, \theta_1)) = (u(a_{\iota(1)}, \theta_2), \ldots, u(a_{\iota(|A|)}, \theta_2))$. Let $\Theta/_e$ be the resulting quotient space whose elements are equivalent classes of exchangeable states.

With ambiguity aversion, a prior homogeneity is no longer sufficient to obtain Luce's model. We need to further require that equivalence classes of exchangeable states are *unambiguous*, i.e., for every $[\theta] \in \Theta/_e$, $\pi([\theta]) = \sum_{\theta \in [\theta]} \pi(\theta)$ is constant across $\pi \in \Pi$. This property is trivially satisfied under expected utility or if all states belong to the same exchangeable class.

Proposition 4. *If problem* (3) *is a priori homogeneous and classes of exchangeable states are unambiguous, then the stochastic choice rule*

$$f(a|\theta) = \frac{\exp[u(a, \theta)/\lambda]}{\sum_{b \in A} \exp[u(b, \theta)/\lambda]} \tag{8}$$

solves RI with smooth ambiguity.

This result tells us that, under our assumptions, Luce's solution (1959) allows the DM to *perfectly hedge* against ambiguity. When actions are a priori homogeneous and classes of exchangeable states are unambiguous, we show that the stochastic choice rule (8) associates each model with the same expected utility amount. Therefore, ambiguity aversion plays no role since every model is equivalent in expected utility terms. This implies that the value of the problem associated with (8) equals the one obtained under expected utility, which is optimal by Matějka and McKay (2015). By Jensen's inequality, it must also be optimal for any ambiguity aversion level.[19]

As a priori homogeneity is defined in terms of utils, it may hold even when actions are not homogeneous according to the smooth certainty equivalent. The following example applies Proposition 4 to find an analytical solution in one of these case.

[18] In their framework, Matějka and McKay (2015) identify states with the payoff associated with each action, treating two payoff-equivalent states as the same. As payoff-equivalent states may differ for their information value, we study a setting where they are not automatically merged by the state description. We can relax the reduction of states to payoff-vectors in this section by assuming that payoff-equivalent states are unambiguous, i.e., every model assigns them the same probability.

[19] By Jensen's inequality, the value of the problem under ambiguity aversion is lower than under expected utility for any stochastic choice rule.

Example 3. The following payoff structures describe decision problems A and B.

A	θ_1	θ_2
a	2	1
b	1	2

B	θ_1	θ_2	θ_3
a	2	1	0
b	1	2	0

We first illustrate our new condition. In problem A, all the states, θ_1 and θ_2, are exchangeable and, therefore, belong to the same exchangeable class, which must be unambiguous. In problem B, where the state θ_3 is also present, we have two classes of exchangeable state: $[\theta_1] = [\theta_2]$ and $[\theta_3]$. Hence, our requirement is satisfied if $\pi(\theta_3)$ is constant for all models in Π.

Consider now problem A. Let the collection of possible models be $\{\pi_1, \pi_2\}$, where $\pi_1 = (5/6,\, 1/6)$ and $\pi_2 = (1/3,\, 2/3)$, and the prior be $\mu = (1/3,\, 2/3)$. Notice that the smooth certainty equivalent, before processing information, differs for action a and b:

$$V^{(\phi,\mu)}(a) = \phi^{-1}\left(\frac{1}{3}\phi\left(\frac{5}{6}\cdot 2 + \frac{1}{6}\cdot 1\right) + \frac{2}{3}\phi\left(\frac{1}{3}\cdot 2 + \frac{2}{3}\cdot 1\right)\right)$$

$$V^{(\phi,\mu)}(b) = \phi^{-1}\left(\frac{1}{3}\phi\left(\frac{5}{6}\cdot 1 + \frac{1}{6}\cdot 2\right) + \frac{2}{3}\phi\left(\frac{1}{3}\cdot 1 + \frac{2}{3}\cdot 2\right)\right)$$

for any ϕ strictly concave. However, as the predictive distribution is $\rho(\theta_1) = \rho(\theta_2) = 1/2$, the problem is a priori homogeneous. By Proposition 4, its solution follows Luce's (1959) multinomial logit model for any level of ambiguity aversion,

$$f(a|\theta_1) = f(b|\theta_2) = \frac{\exp[2/\lambda]}{\exp[2/\lambda] + \exp[1/\lambda]}.$$

◂

4 Axiomatic Characterization

Consider the perspective of an external analyst who observes the DM's behavior conditional on each state of the world. As the analyst does not observe the agent's beliefs, the state space faced by the DM is unknown. For this reason, we include all possible states in the analyst's data set. The objective is to construct statistics based on the available data, composed solely of stochastic choice rules, to verify if DM's behavior is consistent with rational inattention for some ambiguity aversion level.

We modify section 2 notation to relate to Matějka and McKay's (2015) axiomatization more closely. Let P be a finite set of prizes and assume $|A| \geqslant 3$. We define the state space as all possible action-prize combinations, $\Theta := P^A$. For every $a \in A$, we denote by $\theta(a) \in P$ the prize associated with action a when the state is $\theta \in \Theta$. For every state $\theta \in \Theta$, the analyst has access to a data set composed of the choice frequency $f(\cdot|\theta) \in \Delta(A)$. We characterize the properties this data set must satisfy to allow the construction of statistics to represent a utility for each prize $\hat{v}(\theta(a)) := \hat{u}(a, \theta)$, a probability distribution over actions $\hat{f}^0 \in \Delta(A)$, and a state fixed-effect $\hat{I} : \Theta \to \mathbb{R}_+$ such that equation (4) holds, i.e.,

$$f(a|\theta) = \frac{\hat{f}^0(a) \exp[\hat{v}(\theta(a)) \cdot \hat{I}(\theta)]}{\sum_{b \in A} \hat{f}^0(b) \exp[\hat{v}(\theta(b)) \cdot \hat{I}(\theta)]} \tag{9}$$

for every $a \in A$, $\theta \in \Theta$. Notice that we can rescale $\hat{v}$ to allow for $\lambda > 0$.

Whenever the data set does not match the properties characterizing equation (9), this axiomatic approach allows the analyst to infer that the decision maker's behavior is incompatible with rational inattention for any level of ambiguity aversion. The converse does not hold since the statistics in the representation only need to satisfy equation (9) and may not arise from any optimal information acquisition problem. This caveat also applies to the result in Matějka and McKay (2015).

Matějka and McKay (2015) characterize the solution of RI under EU (6) with two axioms that generalize Luce's (1959) *independence of irrelevant alternatives* (IIA).[20] The following matrices of prizes summarize salient aspects of their axiomatization.

A1	θ_1	θ_2	
a	□	□	
b	△	△	

A2	θ_1	θ_2	
a	□	△	
b	□	△	

Figure 2: Matějka and McKay (2015) axiomatization.

[20] IIA can be stated as follows. If $f(a_m|\theta_1) > 0$, then

$$\frac{f(a_n|\theta_1)}{f(a_m|\theta_1)} = \frac{f(a_\ell|\theta_2)}{f(a_k|\theta_2)}$$

for every $a_k, a_\ell, a_m, a_n \in A$ and $\theta_1, \theta_2 \in \Theta$ such that either (i) $\theta_1(a_n) = \theta_2(a_\ell)$ and $\theta_1(a_m) = \theta_2(a_k)$ or (ii) $\theta_1(a_n) = \theta_1(a_m)$ and $\theta_2(a_\ell) = \theta_2(a_k)$. When action a_n coincides with a_ℓ, and action a_m with k, then Axiom 1 corresponds to (i) and Axiom 2 to (ii).

The symbols □ and △ in the entries represent the prizes of playing actions in every state. The two matrices can be viewed as part of a larger decision problem that, for simplicity, we left unspecified.

Matrix A1 describes a situation where the states θ_1 and θ_2 deliver the same prizes for the actions a and b. We say that the states $\theta_1, \theta_2 \in \Theta$ are *payoff-equivalent conditional on* $a, b \in A$ if $\theta_1(a) = \theta_2(a)$ and $\theta_1(b) = \theta_2(b)$. The first axiom postulates that the *probability ratio* induced by the observed stochastic choice rule, which captures the relative frequency of playing two actions conditional on the same state, is unaffected by payoff-equivalent states.

Axiom 1. *If* $f(b|\theta_1) > 0$, *then*

$$\frac{f(a|\theta_1)}{f(b|\theta_1)} = \frac{f(a|\theta_2)}{f(b|\theta_2)}$$

$\forall \theta_1, \theta_2 \in \Theta$, $a, b \in A$ *such that* θ_1 *and* θ_2 *are payoff-equivalent conditional on* a, b.

Matrix A2 describes the opposite situation, as the actions a and b deliver the same prize in each state θ_1 and θ_2. We say that the actions $a, b \in A$ are *duplicates conditional on* $\theta_1, \theta_2 \in \Theta$ if $\theta_1(a) = \theta_1(b)$ and $\theta_2(a) = \theta_2(b)$. The second axiom postulates that the probability ratio induced by the observed stochastic choice rule is unaffected by duplicate actions.

Axiom 2. *If* $f(b|\theta_1) > 0$, *then*

$$\frac{f(a|\theta_1)}{f(b|\theta_1)} = \frac{f(a|\theta_2)}{f(b|\theta_2)}$$

for every $\theta_1, \theta_2 \in \Theta$, $a, b \in A$ *such that* a *and* b *are duplicates conditional on* θ_1, θ_2.

When the DM is ambiguity averse, duplicates are treated as a single action (Proposition 6), implying that Axiom 2 holds. On the other hand, Axiom 1 is no longer satisfied under ambiguity aversion. Analogously to Example 2, payoff-equivalent states may provide different information about models that ambiguity averse DMs could exploit to hedge against ambiguity. Ideally, we would like Axiom 1 to hold only for information-equivalent states, as in weak IUC, but the information value depends on DM's beliefs, which we do not observe.

We can restate Axiom 1 in terms of two separate assumptions. Axiom 1 says that if two states are payoff-equivalent conditional on a pair of actions, then (1.a)

the corresponding probability ratios are well-defined, and (1.b) coincide. While point (1.b) is not satisfied under ambiguity aversion, (1.a) is. We isolate property (1.a) with the following axiom.

Axiom 1.a. *For every $\theta_1, \theta_2 \in \Theta$ and $a, b \in A$ such that θ_1 and θ_2 are payoff-equivalent conditional on a, b, we have that $f(b|\theta_1) > 0 \implies f(b|\theta_2) > 0$.*

We call an action *positive* if it is played with positive probability in every state. The consequences of Axiom 1.a are substantive, as it implies, together with Axiom 2, that every action, played with positive probability in some states, is positive.

Lemma 3. *If Axioms 1.a and 2 hold, then every action is either never played in any state or positive.*

We next discuss how to generalize (1.b) to account for ambiguity aversion. We do so by noticing that, if the states θ_1 and θ_2 are payoff-equivalent conditional on the actions a and b, we can apply equation (7) to relate the corresponding posterior ratios as follows

$$\left(\frac{\rho(\theta_1|a)}{\rho(\theta_1|b)}\right)^{1/\mathcal{I}_{\theta_1}} = \left(\frac{\rho(\theta_2|a)}{\rho(\theta_2|b)}\right)^{1/\mathcal{I}_{\theta_2}}. \tag{10}$$

That is, under ambiguity aversion, posterior ratios may differ across payoff-equivalent states, but their difference vanishes once we adjust for the reciprocal of each state's information value. Under expected utility, since information values equal one in every state, the two posterior ratios coincide, a property called *invariant likelihood ratio*. Notice, by the definition of predictive posterior, the invariant likelihood ratio property is equivalent to (1.b). Therefore, equation (10) is our candidate to generalize (1.b).

Constructing statistics for equation (10) is not immediate as the analyst does not observe posterior distributions and information values directly. We rely on the following construction. First, as in Matějka and McKay (2015), we require that at least three actions are positive. Let these actions be a_1, a_2 and a_N according to the enumeration $A := \{a_1, a_2, \ldots, a_N\}$. We denote by $\theta^k \in \Theta$ any state that assigns the same prize to actions a_1 and a_k,

$$\theta^k = (\theta(a_k), \theta(a_2), \ldots, \theta(a_k), \ldots, \theta(a_N)).$$

For every action $a_k \in A$ and state $\theta^k \in \Theta$ we use the following probability ratio

$$\xi_k := \frac{f(a_k|\theta^k)}{f(a_1|\theta^k)}$$

as a sufficient statistics for $f^0(a_k)/f^0(a_1)$. That is, ξ_k captures the relative likelihood of selecting action a_k over action a_1 for states that assign the same prize to the two actions. Notice that, for every action $a_k \in A$, ξ_k is well-defined by Lemma 3.

For every payoff-equivalent state $\theta_1, \theta_2 \in \Theta$ conditional on the actions $a_\ell, a_k \in A$, we define the following statistic whenever possible

$$\eta_{\ell,\,k}(\theta_1, \theta_2) := \log\left[\frac{\xi_k \cdot f(a_\ell|\theta_1)}{\xi_\ell \cdot f(a_k|\theta_1)}\right] \Big/ \log\left[\frac{\xi_k \cdot f(a_\ell|\theta_2)}{\xi_\ell \cdot f(a_k|\theta_2)}\right]. \tag{11}$$

For any state $\theta \in \Theta$, the fraction $(\xi_k \cdot f(a_\ell|\theta))/(\xi_\ell \cdot f(a_k|\theta))$ constitutes a statistic for the posterior ratio $\rho(\theta|a_\ell)/\rho(\theta|a_k)$ since ξ_k/ξ_ℓ is a statistic for $f^0(a_k)/f^0(a_\ell)$. Therefore, by comparing equations (10) and (11), we term $\eta_{\ell,\,k}(\theta_1, \theta_2)$ the *relative information value* of state θ_1 compared to state θ_2 when the actions a_ℓ and a_k are considered.

The following axiom ensures we can interpret $\eta_{\ell,\,k}(\theta_1, \theta_2)$ as a statistic for $\mathcal{I}(\theta_1)/\mathcal{I}(\theta_2)$.

Axiom 1.b*. *For every payoff-equivalent state $\theta_1, \theta_2 \in \Theta$ conditional on some actions $a_m, a_n \in A$, if $\eta_{n,\,m}(\theta_1, \theta_2)$ is well-defined, then $\exists\, \eta : \Theta \times \Theta \to \mathbb{R}_+$ such that*

1. $\eta_{n,\,m}(\theta_1, \theta_2) = \eta(\theta_1, \theta_2)$.

Furthermore, for every $\theta_i \in \Theta$ such that $\eta(\theta_1, \theta_i)$ and $\eta(\theta_i, \theta_2)$ are well-defined

2. $\eta(\theta_1, \theta_2) = \eta(\theta_1, \theta_i) \cdot \eta(\theta_i, \theta_2)$, *i.e., it does not depend on the choice of $\theta_i \in \Theta$.*

Point 1 of Axiom 1.b* is a form of *action-pair independence*. It says the relative information value between states θ_1 and θ_2 does not depend on a specific action pair. Point 2 is a form of *transitivity*. It implies that, for every state $\theta_i \in \Theta$, $\eta(\theta_1, \theta_i) \cdot \eta(\theta_i, \theta_2)$ is independent of i and equivalent to $\eta(\theta_1, \theta_2)$, provided that $\eta(\theta_1, \theta_i)$ and $\eta(\theta_i, \theta_2)$ are well-defined.

Axiom 1.b* generalizes the property (1.b) of Axiom 1. Indeed, if $\eta_{\ell,\,k}(\theta_1, \theta_2)$ is well-defined for $a_\ell, a_k \in A$, $\theta_1, \theta_2 \in \Theta$, (1.$b$) implies that $\eta_{\ell,\,k}(\theta_1, \theta_2) = 1$.

We are finally ready to state our representation.

Theorem 2. *Assume that three actions are positive. The following are equivalent:*

(*i*) *Axioms 1.a, 1.b* and 2 hold,*

(*ii*) *There exist* $\hat{v} : P \to \mathbb{R}$, $\hat{I} : \Theta \to \mathbb{R}_+$ *and* $\hat{f}^0 \in \Delta(A)$ *such that*

$$f(a|\theta) = \frac{\hat{f}^0(a)\exp[\hat{v}(\theta(a)) \cdot \hat{I}(\theta)]}{\sum_{b \in A} \hat{f}^0(b)\exp[\hat{v}(\theta(b)) \cdot \hat{I}(\theta)]} \tag{12}$$

for every action $a \in A$ *and state* $\theta \in \Theta$.

By replacing Axiom 1 with the Axioms 1.a and 1.b*, we are able to represent a multiplicative fixed-effect $\hat{I} : \Theta \to \mathbb{R}_+$, providing, in light of equation (4), a behavioral characterization of rational inattention with ambiguity aversion. This result establishes that Axiom 1.b*, and therefore equation (10), suitable generalizes the invariant likelihood ratio to the ambiguity averse case.

5 Additional Results

We conclude by discussing additional results related to our setting.

CARA. In our analysis we impose only minor restrictions (Assumption 1) on ambiguity attitudes. The following result states additional properties that arise when ϕ satisfies constant absolute ambiguity aversion (CARA), i.e., $\phi(x) = -\frac{1}{\gamma}e^{-\gamma x}$, $\gamma \in \mathbb{R}_+$.

Proposition 5. *Let* $\theta \in \Theta$, *and* f *be a stochastic choice rule. If* ϕ *satisfies CARA, the following holds:*

(*i*) $\mathbb{E}_\rho[\mathcal{I}_\theta(f)] = 1$,

(*ii*) $\mathcal{I}_\theta(f) \geqslant 1$ *if and only if* $\phi^{-1}(\mathbb{E}_{\mu_\theta}[\,\phi\left(U(\pi, f)\right)]) \leqslant \phi^{-1}(\mathbb{E}_{\mu}[\,\phi\left(U(\pi, f)\right)])$.

Point (*i*) says that information values average 1. This implies that not all states can simultaneously have high or low information value, a property that seems natural in our context. Point (*ii*) strengthens our interpretation of "bad states" (Proposition 1) by establishing that state $\theta \in \Theta$ has high (low) information value whenever computing the certainty equivalent using μ_θ instead of μ reduces (increases) its value.

Maxmin preferences. We exploit the robustness of Proposition 4 to derive optimal stochastic choice rules under extreme ambiguity aversion. Recall that, if ϕ satisfies CARA, then maxmin expected utility (MMEU) by Gilboa and Schmeidler (1989) is a limiting case of smooth ambiguity

$$\lim_{\gamma \to +\infty} V^{(\phi,\mu)}(a) = \min_{\pi \in \Pi} \sum_{\theta \in \Theta} u(a,\theta)\pi(\theta) \tag{13}$$

for every $a \in A$ and μ with support Π.

Contrary to the smooth model, which allows the definition of a predictive distribution, MMEU does not provide any criterion to aggregate models. To apply the model to our setup, we define the entropy-based cost with respect to probability distributions that satisfy a minimal assumption of coherence. We say that $\psi \in \Delta(\Theta)$ is *generated* by Π if there exists a full-support distribution $\eta \in \Delta(\Pi)$ such that $\mathbb{E}_\eta[\pi] = \psi$.

For some $\varphi \in \Delta(\Theta)$, rational inattention with extreme ambiguity aversion,[21] stated in terms of stochastic choice rules, is defined as

$$\max_f \min_{\pi \in \Pi} \sum_{\theta \in \Theta} \sum_{a \in A} u(a,\theta)\, f(a|\theta)\, \pi(\theta) \; - c(f,\varphi). \tag{14}$$

Corollary 1 strengthens Proposition 4 by stating that the multinomial logit model solves RI with extreme ambiguity aversion if it is possible to generate one predictive probability distribution that makes the problem a priori homogeneous.

Corollary 1. *If problem* (14) *is a priori homogeneous for some* $\psi \in \Delta(\Theta)$ *generated by* Π *and classes of exchangeable states are unambiguous, then the stochastic choice rule*

$$f(a|\theta) = \frac{\exp[u(a,\theta)/\lambda]}{\sum_{b \in A} \exp[u(b,\theta)/\lambda]}$$

solves RI with extreme ambiguity aversion.

The result follows by Proposition 4 and the fact that the prior $\mu \in \Delta(\Pi)$ vanishes in the limit of equation (13) except for its support Π.

Duplicates. Consider Luce's (1959) multinomial logit model

$$\mathbb{P}(\text{action } a \text{ chosen from set } A) = \frac{\exp[u(a)]}{\sum_{b \in A} \exp[u(b)]}. \tag{15}$$

[21] This model is not nested in the one studied by Hansen et al. (2022) as they calculate the entropy-based cost with respect to the worst-case prior.

A well-known concern related to this model is Debreu's (1960) red-bus blue-bus criticism. Suppose the DM is indifferent between the actions in the set $\{a, b\}$, e.g., taking a train or a red bus. Following equation (15), this implies that $\mathbb{P}(a) = \mathbb{P}(b) = 1/2$. Assume now that b', a duplicate of action b, is added to the choice set, e.g., a blue bus that differs from the red one only by its color. Then, by equation (15), $\mathbb{P}(a) = \mathbb{P}(b) = \mathbb{P}(b') = 1/3$. This, Debreu argues, is counterintuitive, as adding a duplicate of an existing action decreases the chances of choosing the unduplicated ones, e.g., adding the blue bus decreases the probability of choosing the train.

Matějka and McKay (2015) show that this counterintuitive behavior does not arise under RI. As the duplicates b and b' are a priori equivalent, optimal information processing treats them as a single action. This is a consequence of the role that $f^0 \in \Delta(A)$, which is absent in (15), plays in equation (6), capturing the extent to which each action is a priori appealing.

We now show that the same is true with ambiguity aversion. We say that two actions are duplicates if they yield the same payoff state-wise, i.e., actions $a, b \in A$ are *duplicates* if $u(a, \theta) = u(b, \theta)$ for every $\theta \in \Theta$. Given any RI problem, we can construct the *duplicate problem* by removing action $b \in A$ and substituting it with two duplicates b_1, b_2. Denote the new action set by $\bar{A}$.

Proposition 6. *If f solves RI with smooth ambiguity, then $\bar{f}$ solves the duplicate problem if and only if*

$$f(a|\theta) = \bar{f}(a|\theta)$$
$$f(b|\theta) = \bar{f}(b_1|\theta) + \bar{f}(b_2|\theta)$$

for every $a \in \bar{A} \backslash \{b_1, b_2\}$. In particular, $\bar{f}(b_1|\theta) = f(b|\theta)$ and $\bar{f}(b_2|\theta) = 0$ is a solution.

Proposition 6 implies that adding a duplicate action does not affect the information value of any state. Intuitively, as adding duplicates does not affect the obtainable payoff in each state, it also does not modify states' information value.

Comparative information values. We compare states featuring different information values, and, in light of equation (7), we show that Proposition 2 holds comparatively across them.

Corollary 2. *For every $\theta_1, \theta_2 \in \Theta$, and $a, b \in A$, the following holds:*

$$\frac{\rho(\theta_1|a)}{\rho(\theta_1|b)} \geqslant \frac{\rho(\theta_2|a)}{\rho(\theta_2|b)} \iff \frac{\mathcal{I}_{\theta_1}(f)}{\mathcal{I}_{\theta_2}(f)} \geqslant \frac{u(a, \theta_2) - u(b, \theta_2)}{u(a, \theta_1) - u(b, \theta_1)}.$$

The equivalence above says that the "recommendation" to play $a \in A$ is more informative than $b \in A$ about state $\theta_1 \in \Theta$ compared to state $\theta_2 \in \Theta$ whenever the ratio of the corresponding information values $\mathcal{I}_{\theta_1}/\mathcal{I}_{\theta_2}$ is greater than the ratio of the rewards of distinguishing action a from b in both states.

In particular, Corollary 2 implies that whenever the payoff incentives are held constant across the two states, i.e., $u(a,\theta_1) - u(b,\theta_1) = u(a,\theta_2) - u(b,\theta_2)$, then state $\theta_1 \in \Theta$ is of *higher information value* than state $\theta_2 \in \Theta$, which we define as $\mathcal{I}_{\theta_1}(f) \geqslant \mathcal{I}_{\theta_2}(f)$, if and only if the corresponding posterior ratio is also higher. If the rewards of distinguishing action a from b in states θ_1 and θ_2 are equal, then any attentional difference across the two states must be attributed to their information values.

Information on models. We investigate the consequences of altering the information processing domain by allowing the DM to pay attention to models instead of payoff-relevant states. We develop a possible framework that can be re-interpreted as the standard rational inattention problem, where the properties discussed by Matějka and McKay (2015), among others, still hold.

The information structure $G : \Pi \to \Delta(X)$, mapping each model $\pi \in \Pi$, where Π is assumed to be finite, to a signal distribution in $\Delta(X)$, represents the DM's attention allocation. The entropy-based information cost takes the following form

$$k(\mu, G) := \lambda(\tilde{H}(\mu) - \mathbb{E}_x[\tilde{H}(\mu(\cdot|x))]),$$

where $\lambda \geqslant 0$, $\tilde{H} : \Delta(\Pi) \to \mathbb{R}_+$ is the entropy function defined over the set of priors, and $\mu(\cdot|x) \in \Delta(\Pi)$ is the posterior obtained by the Bayesian updating of the prior μ after observing the signal $x \in X$. The DM solves the following problem

$$\max_{G:\Pi\to\Delta(X)} \max_{\alpha:X\to A} \left(\sum_{\pi\in\Pi} \sum_{x\in X} \phi \left(\sum_{\theta\in\Theta} u(\alpha(x),\theta)\, \pi(\theta) \right) G(x|\pi)\, \mu(\pi) - k(\mu, G) \right) \tag{16}$$

Paralleling assumption (ii) of Definition 1, where information on states does not affect the prior over models, information on models leaves the information on states intact and updates only the prior μ in equation (16). Furthermore, the certainty equivalent in the equation is not expressed in terms of utils via the normalization ϕ^{-1} as in (2), but in ambiguous utils directly.

Re-interpreting equation (16) as a rational inattention problem under expected utility is relatively straightforward. To do so, treat Π as the set of payoff-relevant states of the new problem and set the utility as $\tilde{u}(a,\pi) := \phi(\sum_{\Theta} u(\alpha(x),\theta)\,\pi(\theta))$ for every $a \in A$, $\pi \in \Pi$. The prior over models μ plays the role of prior over payoff-relevant states, and the information cost function k is unchanged.

If $\lambda > 0$, the stochastic choice rule $g : \Pi \to \Delta(A)$ that solves (16) satisfies

$$g(a|\pi) = \frac{g^0(a)\exp[\tilde{u}(a,\pi)/\lambda]}{\sum_{b\in A} g^0(b)\exp[\tilde{u}(b,\pi)/\lambda]}$$

and the consumer optimally trade-off utility gains associated with each model $\tilde{u}(a,\pi)$ and information costs λ. Interestingly, under this re-labeling, ambiguity attitudes ϕ play the same role that risks attitudes $\tilde{u}$ play under expected utility.

Appendix

Proof of Lemma 1. We proceed in four steps.

Step 1. For every strategy (F,α), we can rewrite the certainty equivalent in equation (2) as follows

$$\begin{aligned}
&\phi^{-1}(\mathbb{E}_\mu[\phi(\textstyle\sum_{\theta\in\Theta}\sum_{x\in X} u(\alpha(x),\theta)\,F(x|\theta)\,\pi(\theta))]) \\
&= \phi^{-1}(\mathbb{E}_\mu[\phi(\textstyle\sum_{\theta\in\Theta}\sum_{a\in A}\sum_{x_a\in X_a} u(\alpha(x_a),\theta)\,F(x_a|\theta)\,\pi(\theta))]) \\
&= \phi^{-1}(\mathbb{E}_\mu[\phi(\textstyle\sum_{\theta\in\Theta}\sum_{a\in A} u(a,\theta)\sum_{x_a\in X_a} F(x_a|\theta)\,\pi(\theta))]) \\
&= \phi^{-1}(\mathbb{E}_\mu[\phi(\textstyle\sum_{\theta\in\Theta}\sum_{a\in A} u(a,\theta) f(a|\theta)\,\pi(\theta))]) = \phi^{-1}(\mathbb{E}_\mu[\phi(U(\pi,f))]).
\end{aligned}$$

where f is the stochastic choice rule generated by (F,α).

Step 2. We prove that any optimal strategy implies a constant posterior predictive distribution on each X_a, for $a \in A$. We do so by showing that any optimal strategy (F,α) that induces $\rho(\cdot|x') \neq \rho(\cdot|x'')$, for $x', x'' \in X_a$, cannot be optimal. We proceed by contradiction. The statement above implies that $F(x'|\theta) \neq F(x''|\theta)$ for some $\theta \in \Theta$. We construct a new strategy $(\hat{F} : \Theta \to \Delta(\hat{X}),\, \hat{\alpha} : \hat{X} \to A)$ as follows: $\hat{X} := (X \backslash \{x', x''\}) \cup \{\hat{x}\}$; $\hat{F}$ equals F except for $\hat{F}(\hat{x}|\theta) = F(x'|\theta) + F(x''|\theta)$ for every $\theta \in \Theta$; $\hat{\alpha}$ equals α except for $\hat{\alpha}(\hat{x}) = a$. Notice that, by the concavity of the entropy function, the information structure $\hat{F}$ is less expensive than F. Furthermore, the certainty equivalent associated with $(\hat{F},\, \hat{\alpha})$ is equivalent to the one of (F,α). Indeed, by noticing that F and $\hat{F}$ generate the same stochastic choice rule, the conclusion follows by step 1. Therefore, (F,α) cannot be optimal.

Step 3. We show that $c(\rho, F) = c(f, \rho)$, where F is part of the optimal strategy (F, α) and f is the generated stochastic choice rule. By step 2, if (F, α) is optimal, then predictive posteriors are constant on X_a, i.e., $\rho(\cdot|X_a)$ is constant. Therefore, the joint distribution between states and signals is the same as between actions and states. The equality $c(\rho, F) = c(f, \rho)$ holds by the symmetry of the mutual information.

Step 4. By contradiction, suppose the stochastic choice rule f generated by the strategy σ that solves problem (2) does not solve (3). Assume instead that f' is a solution of (3). Clearly, $V(f') - c(f', \rho) > V(f) - c(f, \rho)$. By steps 1 to 3, the optimality of σ implies that f yields the same value as σ. However, as every stochastic choice rule and its induced strategy yield the same value by construction, the strategy σ' induced by f' yields a higher value than σ. Therefore, σ cannot solve problem (2). The other direction is similarly shown. $\square$

Proof of Lemma 2. Under Assumption 1, Proposition 1 in Hennessy and Lapan (2006), which builds on Hardy et al. (1934) Theorem 106.(i), implies that the certainty equivalent V is concave in U when Π is finite. The proof for infinite Π follows the exact same steps and is thus omitted.

For every $\theta \in \Theta$, consider the stochastic choice rule $f(\cdot|\theta) = \beta g(\cdot|\theta) + (1-\beta)h(\cdot|\theta)$, where $\beta \in [0, 1]$, obtained by convex combination of the stochastic choice rules g and h. We have that

$$
\begin{aligned}
&\phi^{-1}(\mathbb{E}_\mu[\phi(U(\pi, f))]) = \phi^{-1}(\mathbb{E}_\mu[\phi(\textstyle\sum_{\theta\in\Theta}\sum_{a\in A} u(a, \theta) f(a|\theta)\,\pi(\theta))]) \\
&= \phi^{-1}(\mathbb{E}_\mu[\phi(\textstyle\sum_{\theta\in\Theta}\sum_{a\in A} u(a, \theta)(\beta g(a|\theta) + (1-\beta)h(a|\theta))\,\pi(\theta))]) \\
&= \phi^{-1}(\mathbb{E}_\mu[\phi(\beta \textstyle\sum_{\theta\in\Theta}\sum_{a\in A} u(a, \theta) g(a|\theta)\,\pi(\theta) + (1-\beta)\sum_{\theta\in\Theta}\sum_{a\in A} u(a, \theta) h(a|\theta)\,\pi(\theta))]) \\
&= \phi^{-1}(\mathbb{E}_\mu[\phi(\beta U(\pi, g) + (1-\beta) U(\pi, h)]) \\
&\geqslant \beta\phi^{-1}(\mathbb{E}_\mu[\phi(U(\pi, g))]) + (1-\beta)\phi^{-1}(\mathbb{E}_\mu[\phi(U(\pi, h))])
\end{aligned}
$$

where the last inequality follows by concavity of V with respect to U. Therefore, the value function of problem (3) is concave as it is the sum of concave functions.

Furthermore, the domain of problem (3), $\Delta(A)^\Theta$, is convex with respect to the convex combination of stochastic choice rules defined above. $\square$

Proof of Theorem 1. The Lagrangian associated with problem (3) is

$$\phi^{-1}(\mathbb{E}_\mu[\phi(\textstyle\sum_{\theta\in\Theta}\sum_{a\in A} f(a|\theta)u(a,\theta)\,\pi(\theta))]) - c(f,\rho) \\ + \textstyle\sum_{a\in A}\sum_{\theta\in\Theta}\psi_a(\theta)f(a|\theta) - \sum_{\theta\in\Theta}\xi(\theta)(\sum_{a\in A} f^0(a) - 1)$$

where $\psi_a(\theta) \geqslant 0$ and $\xi(\theta)$ are the multipliers of the non negative and the unitary constraint, respectively. By taking the first order conditions, we obtain

$$\phi'(V(f))^{-1}\cdot(\mathbb{E}_\mu[\phi'(U(\pi,f))\pi(\theta)]\cdot u(a,\theta)) \\ + \lambda(\rho(\theta)\log(f^0(a)) - \rho(\theta)\log(f(a|\theta)) + \psi_a(\theta) - \xi(\theta) = 0.$$

Dividing by $\rho(\theta) > 0$ and rearranging, we have that

$$u(a,\theta)\cdot\phi'(V(f))^{-1}\cdot\mathbb{E}_{\mu_\theta}[\phi'(U(\pi,f))] \\ + \lambda(\log(f^0(a)) - \log(f(a|\theta)) + (\psi_a(\theta) - \xi(\theta))/\rho(\theta) = 0.$$

We show that, if $f(a) > 0$, then $f(a|\theta) > 0$, and hence $\psi_a(\theta) = 0$. By contradiction, assume that $f(a|\theta) = 0$. This implies that $\log(f(a|\theta)) = -\infty$. As $\psi_a(\theta) \geqslant 0$, we must have $\xi(\theta) = +\infty$ for the FOC to hold. This implies that, for every $b \in A$, either $\psi_b(\theta) = +\infty$ or $\log(f(b|\theta)) = -\infty$, that is, $f(b|\theta) = 0$. But, $\psi_b(\theta) > 0$ implies $f(b|\theta) = 0$. The argument above implies that $f(b|\theta) = 0$ for every $b \in A$, which contradicts the feasibility of the problem.

If $f^0(a) > 0$, we can rewrite the FOC as

$$f^0(a)\cdot\exp[(u(a,\theta)/\lambda)\cdot\phi'(V(f))^{-1}\cdot\mathbb{E}_{\mu_\theta}[\phi'(U(\pi,f))] - \xi(\theta)/(\lambda\cdot\rho(\theta))] = f(a|\theta)$$

that can be rearranged as equation (4) by summing both sides over $a \in A$ and applying $\sum_{a\in A} f(a|\theta) = 1$. ☐

Proof of Proposition 1. We first show $(i) \iff (ii)$ when the support of μ_θ is a singleton for every $\theta \in \Theta$. Let $\operatorname{supp}\mu_\theta = \pi_\theta$. In this case, $\mathcal{I}_\theta(f) \geqslant 1$ iff $\phi'(U(\pi_\theta, f)) \geqslant \phi'(V(f))$. By concavity of ϕ, ϕ' is decreasing which yields the desired conclusion.

Assume that ϕ is prudent to show $(ii) \implies (i)$. As ϕ is concave, we have that $\mathbb{E}_{\mu_\theta}[U(\pi,f)] \leqslant V(f)$ implies $\phi'(\mathbb{E}_{\mu_\theta}[U(\pi,f)]) \geqslant \phi'(V(f))$. As ϕ' is convex, by Jensen's inequality, $\mathbb{E}_{\mu_\theta}[\phi'(U(\pi,f))] \geqslant \phi'(\mathbb{E}_{\mu_\theta}[U(\pi,f)])$, which concludes.

Assume that ϕ is imprudent to show $(i) \implies (ii)$. As ϕ is concave, we have that $\mathbb{E}_{\mu_\theta}[U(\pi,f)] > V(f)$ implies $\phi'(\mathbb{E}_{\mu_\theta}[U(\pi,f)]) < \phi'(V(f))$. As ϕ' is concave, by Jensen's inequality, $\mathbb{E}_{\mu_\theta}[\phi'(U(\pi,f))] \leqslant \phi'(\mathbb{E}_{\mu_\theta}[U(\pi,f)])$, concluding the proof. ☐

Proof of Proposition 2. By Theorem 1, we have that

$$\frac{\rho^{\phi}(\theta|a)}{\rho^{\phi}(\theta|b)} = \frac{f^{\phi}(a|\theta)\, f^{0,\phi}(b)}{f^{\phi}(b|\theta)\, f^{0,\phi}(a)} = \left(\frac{\exp[u(a,\theta)/\lambda]}{\exp[u(b,\theta)/\lambda]}\right)^{\mathcal{I}_{\theta}(f^{\phi})}.$$

By $u(a,\theta)/\lambda \geqslant u(b,\theta)/\lambda$ and $\mathcal{I}_{\theta}(f^{\phi}) \geqslant 1$, the conclusion immediately follows. □

Proof of Proposition 3. Let θ_1 and θ_2 be payoff and information-equivalent. Notice that, by equation (5), if $\theta_1, \theta_2 \in \Theta$ are information-equivalent, that is $\mu_{\theta_1} = \mu_{\theta_2}$, then $\mathcal{I}_{\theta_1} = \mathcal{I}_{\theta_2}$. For every $a \in A$, we have that

$$\frac{f^0(a)\exp[u(a,\theta_1)\cdot \mathcal{I}_{\theta_1}(f)/\lambda]}{\sum_{b\in A} f^0(b)\exp[u(b,\theta_1)\cdot \mathcal{I}_{\theta_1}(f)/\lambda]} = \frac{f^0(a)\exp[u(a,\theta_2)\cdot \mathcal{I}_{\theta_2}(f)/\lambda]}{\sum_{b\in A} f^0(b)\exp[u(b,\theta_2)\cdot \mathcal{I}_{\theta_2}(f)/\lambda]}.$$

Theorem 1 implies $f(a|\theta_1) = f(a|\theta_2)$ for every $a \in A$, proving the statement. □

Proof of Proposition 4. Fix (ϕ, μ) representing ambiguity aversion, and let φ be a strictly increasing and concave function such that $\varphi \circ \phi = \phi'$. By Jensen's inequality, for every stochastic choice rule f we have that

$$\begin{aligned} V^{(\phi,\mu)}(f) &= \phi^{-1}\Big(\mathbb{E}_{\mu}[\phi(U(\pi,f))]\Big) = (\phi^{-1}\circ\varphi^{-1})\circ\varphi\Big(\mathbb{E}_{\mu}[\phi(U(\pi,f))]\Big) \\ &= \phi'^{-1}\circ\varphi\Big(\mathbb{E}_{\mu}[\phi(U(\pi,f))]\Big) \geqslant \phi'^{-1}\Big(\mathbb{E}_{\mu}[(\varphi\circ\phi)(U(\pi,f))]\Big) \\ &= \phi'^{-1}\Big(\mathbb{E}_{\mu}[\phi'(U(\pi,f))]\Big) = V^{(\phi',\mu)}(f) \end{aligned}$$

This means that the higher ambiguity aversion, the lower the value of the certainty equivalent.

Furthermore, as the certainty equivalents $V^{(\phi,\mu)}$ and $V^{(\phi',\mu)}$ share the same μ, they also induce the same predictive predictive distribution ρ. Hence, they give rise to the same entropy-based costs $c(f,\rho)$ for every f. This implies that

$$\max_f V^{(\phi,\mu)}(f) - c(f,\rho) \geqslant \max_f V^{(\phi',\mu)}(f) - c(f,\rho).$$

Under a priori homogeneity, the RI solution under expected utility follows the multinomial logit model of equation (8) (Proposition 1; Matějka and McKay, 2015). In particular, this formula equates $f(a|\theta_1) = f(b|\theta_2)$ for every pair $(a,\theta_1), (b,\theta_2) \in A \times \Theta$ such that $u(a,\theta_1) = u(b,\theta_2)$ and θ_1, θ_2 are exchangeable. Notably, this implies that

$\sum_{a\in A} f(a|\theta)u(a,\theta)$ is constant for every $\theta \in [\theta] \in \Theta/_e$. We denote the expected utilities associated with each equivalence class $[\theta]$ as $U(\delta_{[\theta]}, f)$.

For every f, we have that

$$U(\pi, f) = \sum_{\theta\in\Theta} \pi(\theta) \sum_{a\in A} f(a|\theta)u(a,\theta) = \sum_{[\theta]\in\Theta/_e} \pi([\theta])\, U(\delta_{[\theta]}, f).$$

As exchangeable classes are unambiguous, for every $[\theta] \in \Theta/_e$ and $\pi \in \Pi$, $\pi([\theta])$ is constant. Hence, $U(\pi, f) = U(f)$ is constant as well for every $\pi \in \Pi$. Furthermore, for every ϕ, we have that

$$V^{(\phi,\mu)}(f) = \phi^{-1}\Big(\mathbb{E}_\mu[\phi(U(\pi, f))]\Big) = \phi^{-1}\Big(\mathbb{E}_\mu[\phi(U(f))]\Big) = U(f).$$

This implies that $V^{(\phi,\mu)}(f)$ is constant in ϕ and equates the certainty equivalent of the problem under expected utility. Hence, f yields the same value as under expected utility and, by Jensen's inequality, is optimal for every ϕ. □

Proof of Theorem 2. The fact that (ii) implies (i) follows directly from equation (12) and the discussion therein. In the remainder, we show that (i) implies (ii).

We introduce additional notation to ease the exposition. We write $\theta_1 \sim^{a,b} \theta_2$ when the states $\theta_1, \theta_2 \in \Theta$ are payoff-equivalent conditional on $a, b \in A$. Similarly, we write $a \sim^{\theta_1,\theta_2} b$ when the actions $a, b \in A$ are duplicates conditional on $\theta_1, \theta_2 \in \Theta$. Let $A := \{a_1, a_2, a_3, \ldots, a_N\}$ with $|A| \geqslant 3$. Recall that $\theta(a_i) \in P$ denotes the prize associated to action $a_i \in A$ by state $\theta \in \Theta := P^A$. For every state θ and prize $p \in P$, denote by $(p, \theta_{-j}) \in \Theta$ the state that associates the prize $\theta(a_i)$ with each action $a_i \in A\backslash\{a_j\}$, and the prize p with action $a_j \in A$. Furthermore, we write $\theta^i = (\theta(a_i), \theta_{-1})$ to denote the state assigning the prizes associated to state θ for every action except at a_1 where it assigns the same prize as action a_i. Similarly, let $\theta^{-i} = (\theta(a_1), \theta_{-i})$ be the state assigning the prizes associated to state θ for every action except at a_i where it assigns same prize as action a_1. Clearly, $\theta^1 = \theta^{-1} = \theta$.

Proof of Lemma 3. The following matrix of prizes, which may be part of a larger decision problem left unspecified, exemplifies the argument below and can be used as a reference.

Without loss of generality, assume that $f(a_1|\theta_1) > 0$ where $\theta_1 \in \Theta$. We want to show that $f(a_1|\theta) > 0$ for every $\theta \in \Theta$. Consider the state θ_1^{-2}. Clearly, $\theta_1 \sim^{a_1,a_3} \theta_1^{-2}$.

	θ_1	θ_1^{-2}	θ_2	θ_3	θ_3^{-2}	
a_1	$\square$	$\square$	$\diamondsuit$	$\bowtie$	$\bowtie$	
a_2	$\triangle$	$\square$	$\diamondsuit$	$\circledast$	$\bowtie$	
a_3	$\bigcirc$	$\bigcirc$	$\heartsuit$	$\bigtriangledown$	$\bigtriangledown$	

Therefore, by Axiom 1.a, we have that $f(a_1|\theta_1^{-2}) > 0$. Now consider a generic state whose prizes associated to action a_1 and a_2 coincide, and call this state $\theta_2 \in \Theta$. By Axiom 2, we have that

$$\frac{f(a_2|\theta_1^{-2})}{f(a_1|\theta_1^{-2})} = \frac{f(a_2|\theta_2)}{f(a_1|\theta_2)}$$

which implies that $f(a_1|\theta_2) > 0$. Finally, consider a generic state and name it $\theta_3 \in \Theta$. By the previous argument, $f(a_1|\theta_3^{-2}) > 0$. Hence, as $\theta_3 \sim^{a_1,a_3} \theta_3^{-2}$, by Axiom 1.a, $f(a_1|\theta_3) > 0$, concluding the argument.

Let $f(a|\theta') = 0$ for some $a \in A$ and $\theta' \in \Theta$. We want to show that $f(a|\theta) = 0$ for all $\theta \in \Theta$. Assume not, that is, there exists $\theta'' \in \Theta$ such that $f(a|\theta'') > 0$. By the previous argument, $f(a|\theta') > 0$, which is a contradiction. $\square$

We proceed with the proof of Theorem 2. We divide the argument in two steps.

Step 1. Assume there exists a function $\hat{I} : \Theta \to \mathbb{R}_+$ such that, for every state $\theta_1, \theta_2 \in \Theta$ with $\theta_1 \sim^{a_\ell,a_k} \theta_2$ for some $a_\ell, a_k \in A$, the following holds

$$\left[\frac{\xi_k \cdot f(a_\ell|\theta_1)}{\xi_\ell \cdot f(a_k|\theta_1)}\right]^{1/\hat{I}(\theta_1)} = \left[\frac{\xi_k \cdot f(a_\ell|\theta_2)}{\xi_\ell \cdot f(a_k|\theta_2)}\right]^{1/\hat{I}(\theta_2)} \tag{17}$$

whenever the corresponding ratios are well-defined. Furthermore, without loss of generality, assume that the actions $a_1, a_2, a_N \in A$ are positive.

Fix a state $\bar{\theta} \in \Theta$ as a reference. Define $\hat{v} : P \to \mathbb{R}$,

$$\hat{v}(p) = \log\left(\xi_N \frac{f(a_1|(p, \bar{\theta}_{-1}))}{f(a_N|(p, \bar{\theta}_{-1}))}\right) \cdot \frac{1}{\hat{I}(p, \bar{\theta}_{-1})}.$$

In this step, we show that $\hat{f}^0(a_i) := \xi_i / \sum_j \xi_j$ for every $a_i \in A$, $\hat{I}$ satisfying equation (17), and $\hat{v}$ defined as above suffice to represent equation (12). If action $a_i \in A$ is never played in any state, then $\xi_i = 0$ and (12) trivially holds. Suppose a_i is positive. For every state $\theta \in \Theta$, we have that

$$1 = \sum_{a_j \in A} f(a_j|\theta) = f(a_i|\theta) \cdot \sum_{a_j \in A} \frac{f(a_j|\theta)}{f(a_i|\theta)} = f(a_i|\theta) \cdot \sum_{a_j \in A} \frac{f(a_j|\theta)/f(a_N|\theta)}{f(a_i|\theta)/f(a_N|\theta)}$$

which implies

$$f(a_i|\theta) = \frac{f(a_i|\theta)/f(a_N|\theta)}{\sum_{a_j \in A} f(a_j|\theta)/f(a_N|\theta)}. \tag{18}$$

As $\theta \sim^{a_i,a_N} \theta^i$ for every $i \in \{1,\ldots,N\}$, we can apply equation (17) to obtain

$$\frac{f(a_i|\theta)}{f(a_N|\theta)} = \frac{\xi_i}{\xi_N} \cdot \left[\frac{\xi_N \cdot f(a_i|\theta^i)}{\xi_i \cdot f(a_N|\theta^i)}\right]^{\hat{I}(\theta)/\hat{I}(\theta^i)}$$

which allows us to rewrite equation (18) as

$$f(a_i|\theta) = \frac{\frac{\xi_i}{\xi_N} \cdot \left[\frac{\xi_N \cdot f(a_i|\theta^i)}{\xi_i \cdot f(a_N|\theta^i)}\right]^{\hat{I}(\theta)/\hat{I}(\theta^i)}}{\sum_{a_j \in A} \frac{\xi_j}{\xi_N} \cdot \left[\frac{\xi_N \cdot f(a_j|\theta^j)}{\xi_j \cdot f(a_N|\theta^j)}\right]^{\hat{I}(\theta)/\hat{I}(\theta^j)}} = \frac{\frac{\xi_i}{\xi_N} \cdot \left[\frac{\xi_N \cdot f(a_1|\theta^i)}{f(a_N|\theta^i)}\right]^{\hat{I}(\theta)/\hat{I}(\theta^i)}}{\sum_{a_j \in A} \frac{\xi_j}{\xi_N} \cdot \left[\frac{\xi_N \cdot f(a_1|\theta^j)}{f(a_N|\theta^j)}\right]^{\hat{I}(\theta)/\hat{I}(\theta^j)}}, \tag{19}$$

where the second equality follows from the definition of ξ_j for $j \in \{1,\ldots,N\}$.

Let $\hat{\theta} \in \Theta$ be a state that assigns to action a_N the same prize of the reference state $\bar{\theta}$, i.e., $\hat{\theta}(a_N) = \bar{\theta}(a_N)$. In this case, as $\hat{\theta}^j \sim^{a_1,a_N} (\hat{\theta}(a_j), \bar{\theta}_{-1}) =: \bar{\theta}^{a_j}$ for every $j \in \{1,\ldots,N\}$, we apply equation (17) to obtain

$$\frac{f(a_1|\hat{\theta}^j)}{f(a_N|\hat{\theta}^j)} = \frac{\xi_1}{\xi_N} \cdot \left[\frac{\xi_N \cdot f(a_1|\bar{\theta}^{a_j})}{\xi_1 \cdot f(a_N|\bar{\theta}^{a_j})}\right]^{\hat{I}(\hat{\theta}^j)/\hat{I}(\bar{\theta}^{a_j})}.$$

By plugging this equation into equation (19), we obtain

$$f(a_i|\hat{\theta}) = \frac{\frac{\xi_i}{\xi_N} \cdot \left[\xi_N \cdot \frac{\xi_1}{\xi_N} \cdot \left[\frac{\xi_N \cdot f(a_1|\bar{\theta}^{a_i})}{\xi_1 \cdot f(a_N|\bar{\theta}^{a_i})}\right]^{\hat{I}(\hat{\theta}^i)/\hat{I}(\bar{\theta}^{a_i})}\right]^{\hat{I}(\hat{\theta})/\hat{I}(\hat{\theta}^i)}}{\sum_{a_j \in A} \frac{\xi_j}{\xi_N} \cdot \left[\xi_N \cdot \frac{\xi_1}{\xi_N} \cdot \left[\frac{\xi_N \cdot f(a_1|\bar{\theta}^{a_j})}{\xi_1 \cdot f(a_N|\bar{\theta}^{a_j})}\right]^{\hat{I}(\hat{\theta}^j)/\hat{I}(\bar{\theta}^{a_j})}\right]^{\hat{I}(\hat{\theta})/\hat{I}(\hat{\theta}^j)}}$$

$$= \frac{\dfrac{\xi_i}{\xi_N} \cdot \xi_1^{\hat{I}(\hat{\theta})/\hat{I}(\hat{\theta}^i)} \cdot \left[\dfrac{\xi_N \cdot f(a_1|\bar{\theta}^{a_i})}{\xi_1 \cdot f(a_N|\bar{\theta}^{a_i})}\right]^{\hat{I}(\hat{\theta})/\hat{I}(\bar{\theta}^{a_i})}}{\sum_{a_j \in A} \dfrac{\xi_j}{\xi_N} \cdot \xi_1^{\hat{I}(\hat{\theta})/\hat{I}(\hat{\theta}^j)} \cdot \left[\dfrac{\xi_N \cdot f(a_1|\bar{\theta}^{a_j})}{\xi_1 \cdot f(a_N|\bar{\theta}^{a_j})}\right]^{\hat{I}(\hat{\theta})/\hat{I}(\bar{\theta}^{a_j})}}$$

$$= \xi_i \cdot \left[\xi_N \frac{f(a_1|\bar{\theta}^{a_i})}{f(a_N|\bar{\theta}^{a_i})}\right]^{\hat{I}(\hat{\theta})/\hat{I}(\bar{\theta}^{a_i})} \Big/ \sum_{a_j \in A} \xi_j \cdot \left[\xi_N \frac{f(a_1|\bar{\theta}^{a_j})}{f(a_N|\bar{\theta}^{a_j})}\right]^{\hat{I}(\hat{\theta})/\hat{I}(\bar{\theta}^{a_j})}$$

where the third inequality follows as $\xi_1 = 1$ by construction. By definition of $\hat{v}$ and $\hat{f}^0(a_i)$, we have

$$f(a_i|\hat{\theta}) = \frac{\hat{f}^0(a_i) \exp[\hat{v}(\hat{\theta}(a_i)) \cdot \hat{I}(\hat{\theta})]}{\sum_{a_j \in A} \hat{f}^0(a_j) \exp[\hat{v}(\hat{\theta}(a_j)) \cdot \hat{I}(\hat{\theta})]}$$

which shows our result for states that share the prize associated with the action a_N with the reference state $\bar{\theta}$.

For every $k \in \{1, N\}$, $\theta, \theta', \theta'' \in \Theta$, $\theta^{\theta'_k, \theta''_N} := (\theta'(a_k), \theta(a_2), \ldots, \theta(a_{N-1}), \theta''(a_N)) \in \Theta$ is the state that assigns each action $a_j \in A \backslash \{a_1, a_N\}$ to the prize $\theta(a_j)$; action a_1 to the prize $\theta'(a_k)$; action a_N to the prize $\theta''(a_N)$. For every $i < N$, as $\theta^N \sim^{a_1, a_i} \theta^{\theta_N, \bar{\theta}_N}$, we can apply equation (17) to obtain

$$f(a_i|\theta^N) = f(a_1|\theta^N) \cdot \frac{\xi_i}{\xi_1} \cdot \left[\frac{\xi_1 \cdot f(a_i|\theta^{\theta_N, \bar{\theta}_N})}{\xi_i \cdot f(a_1|\theta^{\theta_N, \bar{\theta}_N})}\right]^{\frac{\hat{I}(\theta^N)}{\hat{I}(\theta^{\theta_N, \bar{\theta}_N})}} = f(a_1|\theta^N) \cdot \frac{\xi_i}{\xi_1} \cdot \left[\frac{\exp[\hat{v}(\theta(a_i))]}{\exp[\hat{v}(\theta(a_N))]}\right]^{\hat{I}(\theta^N)} \tag{20}$$

where the second equality follows as $\theta^{\theta_N, \bar{\theta}_N}$ share the prize associated with the action a_N with the reference state $\bar{\theta}$. For $i = N$, by Axiom 2, as $a_1 \sim^{\theta^N, \theta^{\bar{\theta}_N, \bar{\theta}_N}} a_N$,

$$f(a_N|\theta^N) = f(a_1|\theta^N) \cdot \frac{f(a_N|\theta^{\bar{\theta}_N, \bar{\theta}_N})}{f(a_1|\theta^{\bar{\theta}_N, \bar{\theta}_N})} = f(a_1|\theta^N) \cdot \frac{\xi_N}{\xi_1} \tag{21}$$

where again the second equality follows for the same reason as above. By plugging equation (21) into equation (20), after performing simple calculations, we obtain

$$f(a_i|\theta^N) = f(a_N|\theta^N) \cdot \frac{\xi_i}{\xi_N} \cdot \left[\frac{\exp[\hat{v}(\theta(a_i))]}{\exp[\hat{v}(\theta(a_N))]}\right]^{\hat{I}(\theta^N)} \tag{22}$$

Since $\sum_{a_i \in A} f(a_i|\theta^N) = 1$, by rearranging, we obtain

$$f(a_N|\theta^N) = \frac{\xi_N \exp[\hat{v}(\theta^N(a_N)) \cdot \hat{I}(\theta^N)]}{\sum_{a_j \in A} \xi_j \exp[\hat{v}(\theta^N(a_j)) \cdot \hat{I}(\theta^N)]} \tag{23}$$

Finally, by plugging equation (23) into equation (22) and rearranging we obtain

$$f(a_i|\theta^N) = \frac{\xi_i \exp[\hat{v}(\theta^N(a_i)) \cdot \hat{I}(\theta^N)]}{\sum_{a_j \in A} \xi_j \exp[\hat{v}(\theta^N(a_j)) \cdot \hat{I}(\theta^N)]} = \frac{\hat{f}^0(a_i) \exp[\hat{v}(\theta^N(a_i)) \cdot \hat{I}(\theta^N)]}{\sum_{a_j \in A} \hat{f}^0(a_j) \exp[\hat{v}(\theta^N(a_j)) \cdot \hat{I}(\theta^N)]} \tag{24}$$

establishing our result for all states that takes the form θ^N.

Let $\theta \in \Theta$ be any state. For every $i < N$, by equation (17), as $\theta \sim^{a_i,a_2} \theta^{\theta_1,\bar{\theta}_N}$, we have that

$$f(a_i|\theta) = f(a_2|\theta) \cdot \frac{\xi_i}{\xi_2} \cdot \left[\frac{\xi_2 \cdot f(a_i|\theta^{\theta_1,\bar{\theta}_N})}{\xi_i \cdot f(a_2|\theta^{\theta_1,\bar{\theta}_N})}\right]^{\frac{\hat{I}(\theta)}{\hat{I}(\theta^{\theta_1,\bar{\theta}_N})}} = f(a_2|\theta) \cdot \frac{\xi_i}{\xi_2} \cdot \left[\frac{\exp[\hat{v}(\theta(a_i))]}{\exp[\hat{v}(\theta(a_2))]}\right]^{\hat{I}(\theta)} \tag{25}$$

where the second equality follows as the statement holds for states that share the prize associated with the action a_N with the reference state $\bar{\theta}$. For $i = N$, by Axiom 2, as $a_2 \sim^{\theta,\theta^N} a_N$, we have that

$$f(a_N|\theta) = f(a_2|\theta) \cdot \frac{\xi_N}{\xi_2} \cdot \left[\frac{\xi_2 \cdot f(a_N|\theta^N)}{\xi_N \cdot f(a_2|\theta^N)}\right]^{\frac{\hat{I}(\theta)}{\hat{I}(\theta^N)}} = f(a_2|\theta) \cdot \frac{\xi_N}{\xi_2} \cdot \left[\frac{\exp[\hat{v}(\theta(a_i))]}{\exp[\hat{v}(\theta(a_2))]}\right]^{\hat{I}(\theta)} \tag{26}$$

where the second equality follows as the statement holds for states that takes the form θ^N. By imposing $\sum_{a_i \in A} f(a_i|\theta) = 1$ in equation (25) and rearranging, we obtain

$$f(a_2|\theta) = \frac{\xi_2 \exp[\hat{v}(\theta(a_2)) \cdot \hat{I}(\theta)]}{\sum_{a_j \in A} \xi_j \exp[\hat{v}(\theta(a_j)) \cdot \hat{I}(\theta)]} \tag{27}$$

By plugging equation (27) into both equation (25) and (26) we obtain

$$f(a_i|\theta) = \frac{\xi_i \exp[\hat{v}(\theta(a_i)) \cdot \hat{I}(\theta)]}{\sum_{a_j \in A} \xi_j \exp[\hat{v}(\theta(a_j)) \cdot \hat{I}(\theta)]} = \frac{\hat{f}^0(a_i) \exp[\hat{v}(\theta(a_i)) \cdot \hat{I}(\theta)]}{\sum_{a_j \in A} \hat{f}^0(a_j) \exp[\hat{v}(\theta(a_j)) \cdot \hat{I}(\theta)]} \tag{28}$$

for every $a_i \in A$, hence showing the statement.

Step 2. We are left to show that, if the Axioms 1.a, 1.b* and 2 hold, we can construct a function $\hat{I} : \Theta \to \mathbb{R}$ such that equation (17) is satisfied.

For every $\theta \in \Theta$, define

$$
\begin{aligned}
P(\theta) &:= \{\theta' \in \Theta : \theta \sim^{a_i,a_j} \theta' \text{ for some } a_i, a_j \in A\} \\
P^2(\theta) &:= \{\theta' \in \Theta : \exists \theta'' \in \Theta,\ \theta'' \in P(\theta') \cap P(\theta)\} \\
P^3(\theta) &:= \{\theta' \in \Theta : \exists \theta'', \theta''' \in \Theta,\ \theta''' \in P(\theta'') \cap P(\theta), \theta'' \in P(\theta')\} \\
&= \{\theta' \in \Theta : \exists \theta'' \in \Theta,\ \theta'' \in P(\theta') \cap P^2(\theta)\}
\end{aligned}
$$

where the last equality follows from the definition of P^2.

For every $\theta \in \Theta$, the correspondence $P(\theta)$ denotes all the states that are payoff-equivalent to θ conditional on some action pair. The correspondence $P^2(\theta)$ maps to the states θ' that may not belong to $P(\theta)$, but are "one step" close to it, that is, there exists a state θ'' which belong to both $P(\theta)$ and $P(\theta')$. Similarly, $P^3(\theta)$ captures the idea of states that are "two steps" close to θ.

It immediately follows that $P^1(\theta) \subseteq P^2(\theta) \subseteq P^3(\theta)$ for every $\theta \in \Theta$. Furthermore $P^3(\theta) = \Theta$, that is, all states are at most two steps close to each other. The argument proceeds as follows. Clearly, $P^3(\theta) \subseteq \Theta$. Now, let $\hat{\theta} \in \Theta$ be any state, we want to show that $\hat{\theta} \in P^3(\theta)$. We have that $\tilde{\theta} := \hat{\theta}^{\theta_1, \hat{\theta}_N} \in P(\hat{\theta})$ since $\tilde{\theta} \sim^{a_2,a_N} \hat{\theta}$. Furthermore, $\theta^* := \tilde{\theta}^{\tilde{\theta}_1, \theta_N} \in P(\tilde{\theta})$ since $\theta^* \sim^{a_1,a_2} \tilde{\theta}$. But also $\theta^* \in P(\theta)$ since $\theta^* \sim^{a_1,a_N} \theta$ by construction, hence proving the claim.

Let $\bar{\theta} \in \Theta$ be the reference state. Let $\hat{I} : \Theta \to \mathbb{R}_+$ be a function such that, for every $\theta \in \Theta \backslash \{\bar{\theta}\}$, $\hat{I}(\theta) = \eta(\theta, \theta')\, \hat{I}(\theta')$, for some $\theta' \in P(\theta)$, and $\hat{I}(\bar{\theta}) = 1$. In what follows we show that such a function is well-defined.

We start by providing an overview of the argument. We need to show that $\hat{I}(\theta) = \eta(\theta, \theta')\, \hat{I}(\theta') = \eta(\theta, \theta'')\, \hat{I}(\theta'')$ for every $\theta', \theta'' \in P(\theta)$. We do so by considering different cases for θ: *(i)* $\theta \in P(\bar{\theta})$; *(ii)* $\theta \in P^2(\bar{\theta}) \backslash P(\bar{\theta})$; *(iii)* $\theta \in P^3(\bar{\theta}) \backslash P^2(\bar{\theta})$. Within each case, we show that $\hat{I}(\theta)$ is well-defined in all regions where θ' may belong to: $P(\bar{\theta})$, $P^2(\bar{\theta}) \backslash P(\bar{\theta})$ or $P^3(\bar{\theta}) \backslash P^2(\bar{\theta})$. This is accomplished by applying Axiom 1.b* point 2 to suitably defined intermediate states $\hat{\theta}$, whose existence is proven for each sub case.

We show the claim for $|A| = 3$. By inspection of the argument, this does not constitute an issue as the the proof follows a similar logic for $|A| \geqslant 4$ and is thus omitted. Intuitively, if the payoff-equivalence relations between $\theta, \theta', \bar{\theta}$ depends on more than 3 different actions, the argument simplifies as it becomes easier to define the intermediate states $\hat{\theta}$ as more equivalences are available.

Consider the following cases. *(i)* Let $\theta \in \Theta$ be such that $\theta \in P(\bar{\theta})$. By definition, $\hat{I}(\theta) = \eta(\theta, \bar{\theta})$. Suppose that $\hat{I}(\theta) = \eta(\theta, \theta_1)\, \hat{I}(\theta_1)$ for $\theta_1 \in P(\bar{\theta}) \cap P(\theta)$. $\hat{I}(\theta)$ is well-defined by Axiom 1.b*. Suppose that $\hat{I}(\theta) = \eta(\theta, \theta_2)\, \hat{I}(\theta_2)$ for $\theta_2 \in (P^2(\bar{\theta}) \backslash P(\bar{\theta})) \cap$

$P(\theta)$. In this case

$$\hat{I}(\theta) = \eta(\theta,\theta_2)\,\hat{I}(\theta_2) = \eta(\theta,\theta_2)\,\eta(\theta_2,\theta^*)\eta(\theta^*,\bar{\theta}) = \eta(\theta,\theta_2)\,\eta(\theta_2,\theta)\eta(\theta,\bar{\theta}) = \eta(\theta,\bar{\theta})$$

where the second equality follows by definition of P^2 for $\theta^* \in P(\theta_2) \cap P(\bar{\theta})$, the third by Axiom 1.b* as $\theta \in P(\theta_2) \cap P(\bar{\theta})$, and fourth by definition of η. The case $\theta_3 \in (P^3(\bar{\theta})\backslash P^2(\bar{\theta})) \cap P(\theta)$ is vacuous as inconsistent with $\theta \in P(\bar{\theta})$. *(ii)* Now, let $\theta \in \Theta$ be such that $\theta \in P^2(\bar{\theta})\backslash P(\bar{\theta})$. By definition, $\hat{I}(\theta) = \eta(\theta,\theta')\hat{I}(\theta')$ for every $\theta' \in P(\theta) \cap P(\bar{\theta})$, which is well-defined by Axiom 1.b*. The case of $\theta_1 \in P(\bar{\theta}) \cap P(\theta)$ is already checked. Suppose that $\hat{I}(\theta) = \eta(\theta,\theta_2)\,\hat{I}(\theta_2)$ for $\theta_2 \in (P^2(\bar{\theta})\backslash P(\bar{\theta})) \cap P(\theta)$. We identify two possible sub cases depending on the choice of $\theta^* \in P(\theta_2) \cap P(\bar{\theta})$. We discuss one of the two and we omit the other, as it follows a similar logic. Consider the following matrix of prizes.

	$\bar{\theta}$	θ	θ_2	θ^*	$\hat{\theta}$
a_1	□	□	□	□	□
a_2	△	▽	▽	△	△
a_3	○	⋈	◇	◇	⋈

Notice that the states in the matrix are consistent with the properties above. We have that

$$\hat{I}(\theta) = \eta(\theta,\theta_2)\,\hat{I}(\theta_2) = \eta(\theta,\theta_2)\,\eta(\theta_2,\theta^*)\eta(\theta^*,\bar{\theta}) = \eta(\theta,\hat{\theta})\,\eta(\hat{\theta},\theta^*)\eta(\theta^*,\bar{\theta}) = \eta(\theta,\hat{\theta})\,\eta(\hat{\theta},\bar{\theta})$$

where the third and fourth equalities follow by Axiom 1.b* as $\hat{\theta} \in P(\theta) \cap P(\theta^*) \cap P(\bar{\theta})$. Suppose that $\hat{I}(\theta) = \eta(\theta,\theta_3)\,\hat{I}(\theta_3)$ for $\theta_3 \in (P^3(\bar{\theta})\backslash P^2(\bar{\theta})) \cap P(\theta)$. The following matrix of prizes discusses only one out of six possible sub cases depending on the choice of $\theta^{**} \in P(\theta^*) \cap P(\bar{\theta})$, where $\theta^* \in P(\theta_3) \cap P(\theta^{**})$ and both θ^* and θ^{**} exists given the assumption about θ_3. The omitted cases follow a similar logic.

	$\bar{\theta}$	θ	θ_3	θ^*	θ^{**}	$\hat{\theta}$
a_1	□	□	♡	♡	♡	□
a_2	△	▽	▽	▽	△	▽
a_3	○	⋈	⋈	○	○	○

Notice that the states in the matrix are consistent with the properties above. We have that

$$\hat{I}(\theta) = \eta(\theta,\theta_3)\,\hat{I}(\theta_3) = \eta(\theta,\theta_3)\,\eta(\theta_3,\theta^*)\eta(\theta^*,\theta^{**})\eta(\theta^{**},\bar{\theta})$$
$$= \eta(\theta,\hat{\theta})\,\eta(\hat{\theta},\theta^*)\eta(\theta^*,\hat{\theta})\eta(\hat{\theta},\bar{\theta}) = \eta(\theta,\hat{\theta})\,\eta(\hat{\theta},\bar{\theta})$$

where the third and fourth equalities follow by Axiom 1.b* as $\hat{\theta} \in P(\theta) \cap P(\theta^*) \cap P(\bar{\theta})$.

(*iii*) Finally, let $\theta \in \Theta$ be such that $\theta \in P^3(\bar{\theta}) \backslash P^2(\bar{\theta})$. By definition, $\hat{I}(\theta) = \eta(\theta,\theta')\hat{I}(\theta')$ for every $\theta' \in P(\theta) \cap P^2(\bar{\theta})$, which is well-defined by Axiom 1.b* and by definition of η. The case $\theta_1 \in P(\bar{\theta}) \cap P(\theta)$ is vacuous as inconsistent with $\theta \in P^3(\bar{\theta}) \backslash P^2(\bar{\theta})$ and the case $\theta_2 \in (P^2(\bar{\theta}) \backslash P(\bar{\theta})) \cap P(\theta)$ is already checked. We are left with $\hat{I}(\theta) = \eta(\theta,\theta_3)\,\hat{I}(\theta_3)$ for $\theta_3 \in (P^3(\bar{\theta}) \backslash P^2(\bar{\theta})) \cap P(\theta)$. Again, the following matrix of prizes discusses only one out of six possible sub cases depending on the choice of $\theta^{**} \in P(\theta^*) \cap P(\bar{\theta})$, where $\theta^* \in P(\theta_3) \cap P(\theta^{**})$ and both θ^* and θ^{**} exists given the assumption about θ_3. The omitted cases follow a similar logic.

	$\bar{\theta}$	θ	θ_3	θ^*	θ^{**}	$\hat{\theta}$	$\tilde{\theta}$
a_1	$\square$	$\heartsuit$	$\heartsuit$	$\square$	$\square$	$\square$	$\heartsuit$
a_2	$\triangle$	$\triangledown$	$\triangledown$	$\triangledown$	$\triangle$	$\triangledown$	$\triangledown$
a_3	$\bigcirc$	$\bowtie$	$\Diamond$	$\Diamond$	$\Diamond$	$\bigcirc$	$\bigcirc$

Notice that the states in the matrix are consistent with the properties above. We have that

$$\hat{I}(\theta) = \eta(\theta,\theta_3)\,\hat{I}(\theta_3) = \eta(\theta,\theta_3)\,\eta(\theta_3,\theta^*)\eta(\theta^*,\theta^{**})\eta(\theta^{**},\bar{\theta})$$
$$= \eta(\theta,\theta_3)\,\eta(\theta_3,\theta^*)\eta(\theta^*,\hat{\theta})\eta(\hat{\theta},\bar{\theta}) = \eta(\theta,\theta_3)\,\eta(\theta_3,\tilde{\theta})\eta(\tilde{\theta},\hat{\theta})\eta(\hat{\theta},\bar{\theta})$$
$$= \eta(\theta,\tilde{\theta})\eta(\tilde{\theta},\hat{\theta})\eta(\hat{\theta},\bar{\theta})$$

where the third equality follows by Axiom 1.b* as $\hat{\theta} \in \cap P(\theta^*) \cap P(\bar{\theta})$, the fourth by Axiom 1.b* as $\tilde{\theta} \in \cap P(\hat{\theta}) \cap P(\theta_3)$, the fifth by Axiom 1.b* as $\tilde{\theta} \in \cap P(\theta)$.

We are left to check whether the function $\hat{I}$ that we have constructed satisfies equation (17). Let $\theta_1, \theta_2 \in \Theta$ with $\theta_1 \sim^{a_\ell,a_k} \theta_2$ for some $a_\ell, a_k \in A$. Hence, $\theta_1 \in P(\theta_2)$. By definition of η, we have that

$$\eta(\theta_1,\theta_2) = \log\left[\frac{\xi_k \cdot f(a_\ell|\theta_1)}{\xi_\ell \cdot f(a_k|\theta_1)}\right] \Big/ \log\left[\frac{\xi_k \cdot f(a_\ell|\theta_2)}{\xi_\ell \cdot f(a_k|\theta_2)}\right].$$

By construction of $\hat{I}$ and the argument made in step 2, we have that $\hat{I}(\theta_1) = \eta(\theta_1,\theta_2)\hat{I}(\theta_2)$ is well-defined. This implies $\eta(\theta_1,\theta_2) = \hat{I}(\theta_1)/\hat{I}(\theta_2)$, which, together with the above definition of η, after rearranging, yields (17). This concludes the proof. $\square$

Proof of Proposition 5. Recall that if ϕ is CARA, i.e., $\phi(x) = -\frac{1}{\gamma}e^{-\gamma x}$ for $\gamma \in \mathbb{R}_+$, we have that $\phi^{-1}(x) = -\frac{1}{\gamma}\log(-\gamma x)$, $\phi'(x) = e^{-\gamma x}$, $(\phi')^{-1}(x) = -\frac{1}{\gamma}\log x$. Using the definitions above, point (i) in the statement follows by definition of $\mathcal{I}_\theta(f)$; point (ii) follows by noticing that $\mathcal{I}_\theta(f) \geqslant 1$ if and only if $(\phi')^{-1}(\mathbb{E}_{\mu_\theta}[\,(\phi')\left(U(\pi,f)\right)]) \leqslant \phi^{-1}(\mathbb{E}_\mu[\,\phi\left(U(\pi,f)\right)])$ as the monotonicity and concavity of ϕ makes $(\phi')^{-1}$ decreasing. □

Proof of Corollary 1. Assume ϕ is CARA with parameter $\gamma > 0$. Notice that, by Proposition 4, under a priori homogeneity and unambiguous classes of exchangeable states the stochastic choice rule satisfying equation (8), call it f^*, is independent of γ and $f^* \in \arg\max_f V^{(\gamma,\mu)}(f) - c(f,\psi)$ for every $\gamma > 0$. Thus,

$$V^{(\gamma,\mu)}(f^*) - c(f^*,\psi) \geqslant V^{(\gamma,\mu)}(f) - c(f,\psi)$$

for every stochastic choice rule f, and $\gamma > 0$. By taking limits, we obtain

$$\lim_{\gamma\to+\infty} V^{(\gamma,\mu)}(f^*) - c(f^*,\psi) \geqslant \lim_{\gamma\to+\infty} V^{(\gamma,\mu)}(f) - c(f,\psi),$$

which implies $f^* \in \arg\max_f \lim_{\gamma\to+\infty} V^{(\gamma,\mu)}(f) - c(f,\psi)$. □

Proof of Proposition 6. For every $\pi \in \Pi$, we have that

$$U(\pi,\bar{f}) = \sum_{\theta\in\Theta}\pi(\theta) \sum_{a\in\bar{A}\backslash\{b_1,b_2\}} \bar{f}(a|\theta)u(a,\theta) + u(b,\theta)\sum_{c\in\{b_1,b_2\}}\bar{f}(c|\theta) = U(\pi,f)$$

since $\bar{f}(a|\theta) = f(a|\theta)$ for every $a \in \bar{A}\backslash\{b_1,b_2\}$ and $\sum_{c\in\{b_1,b_2\}}\bar{f}(c|\theta) = f(b|\theta)$. Thus, the solutions f and $\bar{f}$ are associated with the same certainty equivalent, i.e., $V(f) = V(\bar{f})$. Furthermore, by equation (5), this implies that $\mathcal{I}_\theta(f) = \mathcal{I}_\theta(\bar{f})$ for every $\theta \in \Theta$.

The solutions f and $\bar{f}$ are also associated with the same entropy costs. To see this, notice that for any stochastic choice rule g, the entropy-based cost $c(g,\rho)$ can be written as

$$c(g,\rho) = \sum_{\theta\in\Theta}\rho(\theta)\sum_{a\in A} g(a|\theta)\log\left(\frac{g(a|\theta)}{g^0(a)}\right).$$

Applying Theorem 1, we have that

$$c(\bar{f},\rho) = \sum_{\theta\in\Theta}\rho(\theta)\sum_{a\in\bar{A}}\bar{f}(a|\theta)\log\left(\frac{\exp[u(a,\theta)\,\mathcal{I}_\theta(\bar{f})/\lambda]}{\sum_{a\in\bar{A}}\bar{f}^0(a)\exp[u(a,\theta)\,\mathcal{I}_\theta(\bar{f})/\lambda]}\right) = c(f,\rho),$$

where the last equality follows from $\exp[u(b_i,\theta)\,\mathcal{I}_\theta(\bar{f})/\lambda] = \exp[u(b,\theta)\,\mathcal{I}_\theta(f)/\lambda]$ for $i \in \{1,2\}$ and $f^0(b) = \sum_\theta[\bar{f}^0(b_1|\theta) + \bar{f}^0(b_2|\theta)]\rho(\theta) = \bar{f}^0(b_1) + \bar{f}^0(b_2)$. □

Proof of Corollary 2. By equation (7), we have that

$$\frac{\rho(\theta_1|a)}{\rho(\theta_1|b)} \geqslant \frac{\rho(\theta_2|a)}{\rho(\theta_2|b)} \iff \left(\frac{\exp[u(a,\theta_1)/\lambda]}{\exp[u(b,\theta_1)/\lambda]}\right)^{\mathcal{I}_{\theta_1}(f^\phi)} \geqslant \left(\frac{\exp[u(a,\theta_2)/\lambda]}{\exp[u(b,\theta_2)/\lambda]}\right)^{\mathcal{I}_{\theta_2}(f^\phi)}.$$

Applying a logarithmic transformation on both sides of the inequality on the right yields the desired equivalence. □

Bibliography

Amari, S. and H. Nagaoka (2000): *Methods of information geometry*, vol. 191, American Mathematical Soc.

Angeletos, G.-M. and K. A. Sastry (2024): "Inattentive Economies," Working Paper.

Auster, S., Y.-K. Che, and K. Mierendorff (2024): "Prolonged Learning and Hasty Stopping: the Wald Problem with Ambiguity," *American Economic Review*, forthcoming.

Battigalli, P., S. Cerreia-Vioglio, F. Maccheroni, and M. Marinacci (2015): "Self-confirming equilibrium and model uncertainty," *American Economic Review*, 105, 646–677.

Caplin, A. and M. Dean (2013): "Behavioral implications of rational inattention with shannon entropy," Tech. rep., National Bureau of Economic Research.

Caplin, A., M. Dean, and J. Leahy (2019): "Rational inattention, optimal consideration sets, and stochastic choice," *The Review of Economic Studies*, 86, 1061–1094.

——— (2022): "Rationally inattentive behavior: Characterizing and generalizing Shannon entropy," *Journal of Political Economy*, 130, 1676–1715.

Csiszár, I. (1974): "On an extremum problem of information theory," *Studia Sci. Math. Hungar*, 9, 57–61.

De Oliveira, H., Y. Ishii, and X. Lin (2021): "Robust Merging of Information," *arXiv preprint arXiv:2106.00088.*

Dean, M. and N. Neligh (2023): "Experimental tests of rational inattention," *Journal of Political Economy*, 131, 3415–3461.

Debreu, G. (1960): "Review of RD Luce, Individual choice behavior: A theoretical analysis," *American Economic Review*, 50, 186–188.

Denti, T., M. Marinacci, and L. Montrucchio (2020): "A note on rational inattention and rate distortion theory," *Decisions in Economics and Finance*, 43, 75–89.

Denti, T. and L. Pomatto (2022): "Model and predictive uncertainty: A foundation for smooth ambiguity preferences," *Econometrica*, 90, 551–584.

Epstein, L. G. and S. Ji (2022): "Optimal learning under robustness and time-consistency," *Operations Research*, 70, 1317–1329.

Epstein, L. G. and M. Schneider (2003): "Recursive multiple-priors," *Journal of Economic Theory*, 113, 1–31.

Gilboa, I. and D. Schmeidler (1989): "Maxmin expected utility with non-unique prior," *Journal of Mathematical Economics*, 141–151.

Hanany, E. and P. Klibanoff (2009): "Updating ambiguity averse preferences," *The BE Journal of Theoretical Economics*, 9.

Hansen, L. P., J. Miao, and H. Xing (2022): "Robust rationally inattentive discrete choice," Tech. rep., Working Paper.

Hardy, G. H., J. E. Littlewood, G. Pólya, G. Pólya, et al. (1934): *Inequalities*, Cambridge university press.

Hébert, B. and J. LaO (2023): "Information acquisition, efficiency, and nonfundamental volatility," *Journal of Political Economy*, 131, 2666–2723.

Hennessy, D. A. and H. E. Lapan (2006): "On the nature of certainty equivalent functionals," *Journal of Mathematical Economics*, 43, 1–10.

Klibanoff, P. and E. Hanany (2007): "Updating preferences with multiple priors," *Theoretical Economics*, 2, 261–298.

Klibanoff, P., M. Marinacci, and S. Mukerji (2005): "A smooth model of decision making under ambiguity," *Econometrica*, 73, 1849–1892.

Li, J. (2020): "Preferences for partial information and ambiguity," *Theoretical Economics*, 15, 1059–1094.

Li, J. and J. Zhou (2016): "Blackwell's informativeness ranking with uncertainty-averse preferences," *Games and Economic Behavior*, 96, 18–29.

Lu, J. (2021): "Random ambiguity," *Theoretical Economics*, 16, 539–570.

Luce, R. D. (1959): "Individual choice behavior: A theoretical analysis, new york, ny: John willey and sons," .

MACCHERONI, F., M. MARINACCI, AND A. RUSTICHINI (2006): "Ambiguity aversion, robustness, and the variational representation of preferences," *Econometrica*, 74, 1447–1498.

MAĆKOWIAK, B. AND M. WIEDERHOLT (2009): "Optimal sticky prices under rational inattention," *American Economic Review*, 99, 769–803.

MATĚJKA, F. AND A. MCKAY (2015): "Rational inattention to discrete choices: A new foundation for the multinomial logit model," *American Economic Review*, 105, 272–298.

MELE, A. AND F. SANGIORGI (2015): "Uncertainty, information acquisition, and price swings in asset markets," *The Review of Economic Studies*, 82, 1533–1567.

MONDRIA, J. (2010): "Portfolio choice, attention allocation, and price comovement," *Journal of Economic Theory*, 145, 1837–1864.

SIMS, C. A. (2003): "Implications of rational inattention," *Journal of monetary Economics*, 50, 665–690.

SINISCALCHI, M. (2011): "Dynamic choice under ambiguity," *Theoretical Economics*, 6, 379–421.

VAN NIEUWERBURGH, S. AND L. VELDKAMP (2010): "Information acquisition and under-diversification," *The Review of Economic Studies*, 77, 779–805.